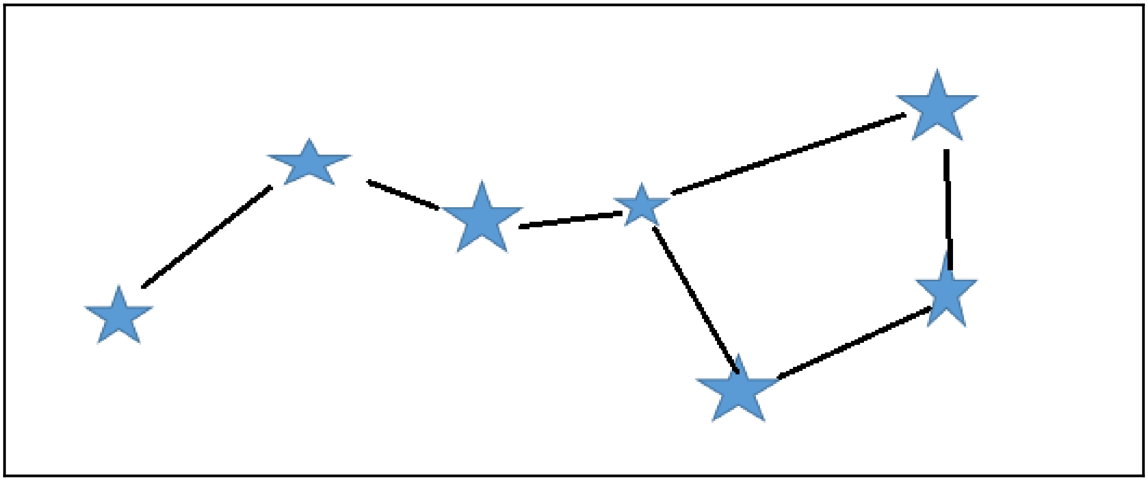
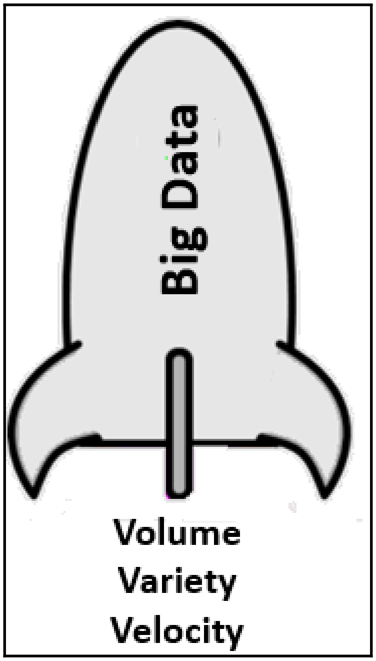
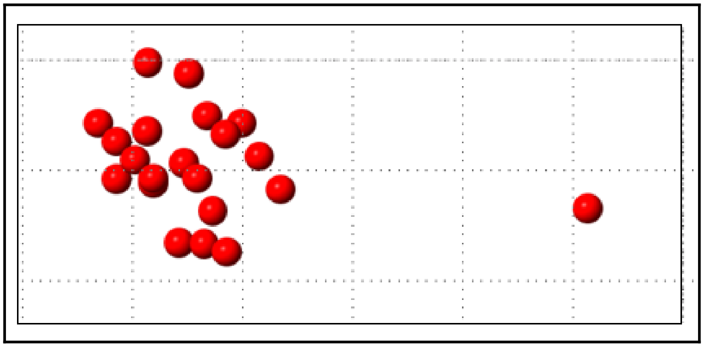
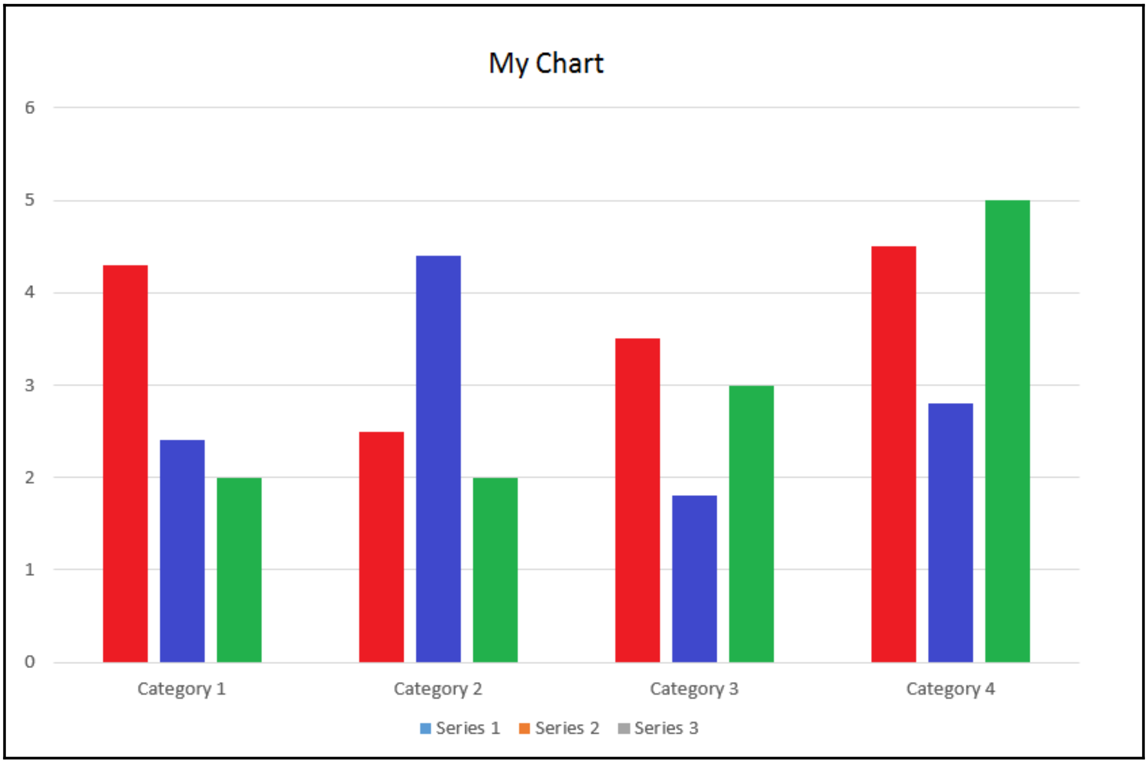
Chapter 1: Introduction to Big Data Visualization

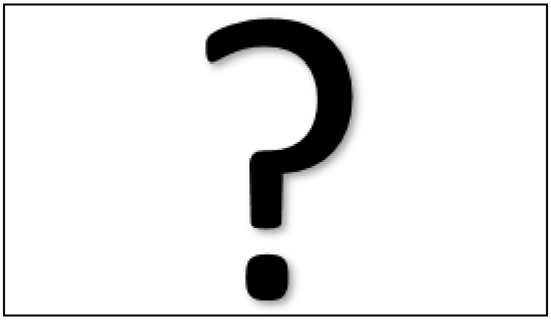



Big Data
Volume
Variety
Velocity

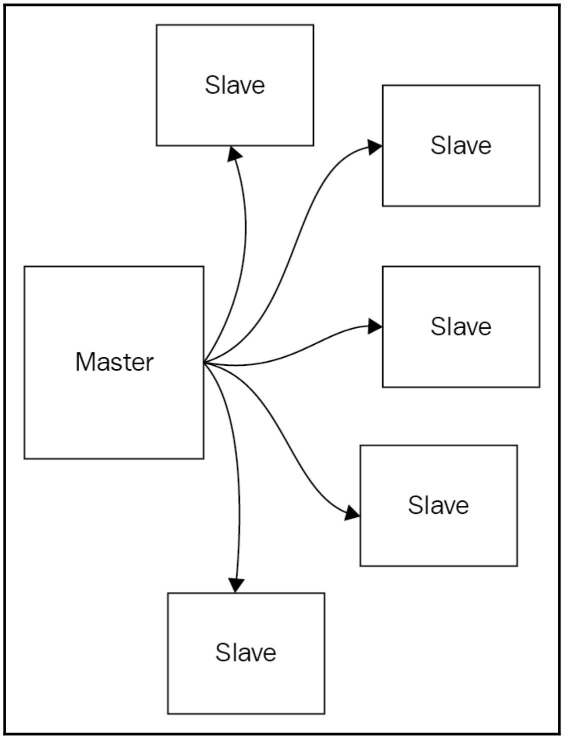
Quality = Value
$
$
$
$
$
$
$
Quality
Value

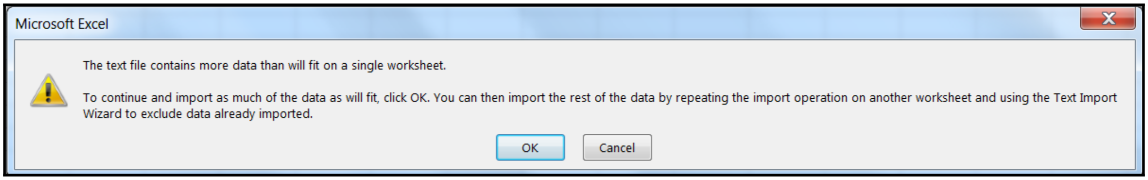



My Chart
6
5
4
3
2
1
0
Category 1
Category 2
Category 3
Category 4
🟦 Series 1
🟧 Series 2
⬜ Series 3



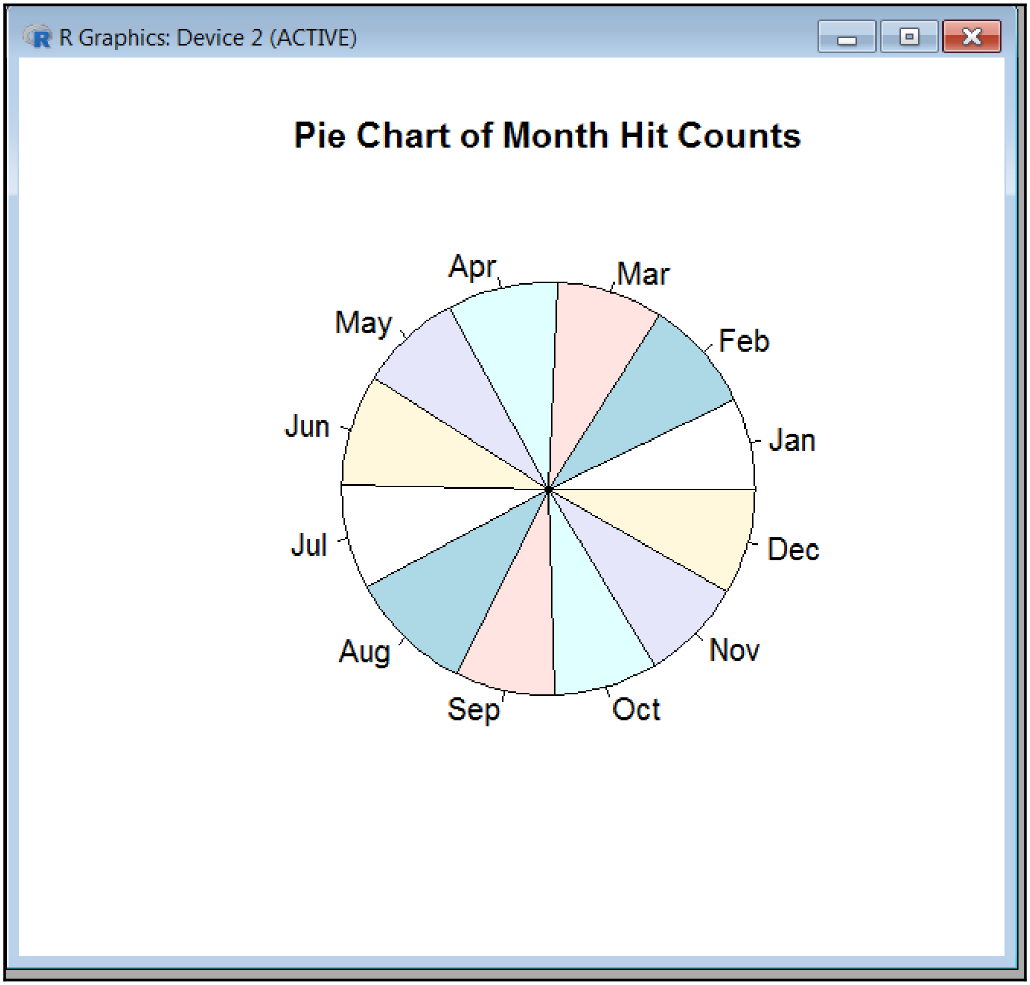
Chapter 2: Access, Speed, and Storage with Hadoop

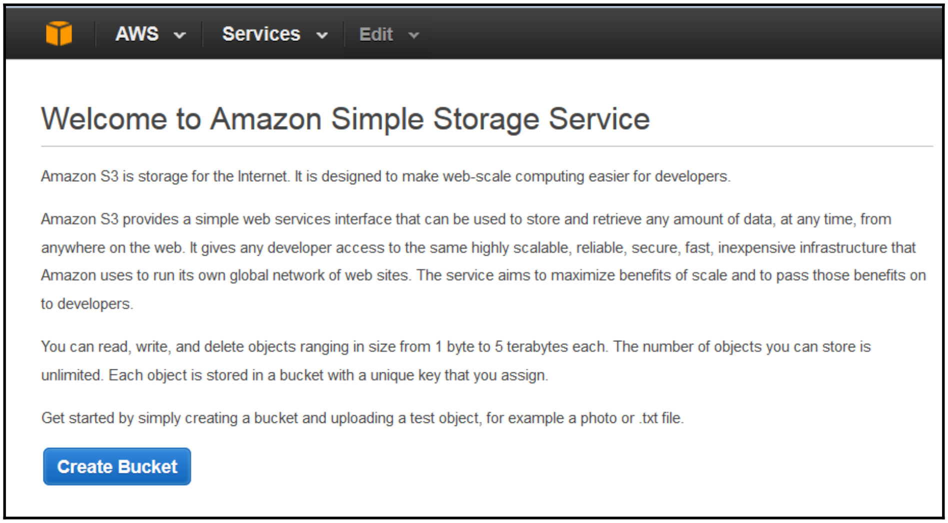




The text file contains more data than will fit on a single worksheet.

To continue and import as much of the data as will fit, click OK. You can then import the rest of the data by repeating the import operation on another worksheet and using the Text Import Wizard to exclude data already imported.


R Graphics: Device 2 (ACTIVE)
Pie Chart of Month Hit Counts
Apr
Mar
May
Feb
Jun
Jan
Jul
Dec
Aug
Nov
Sep
Oct



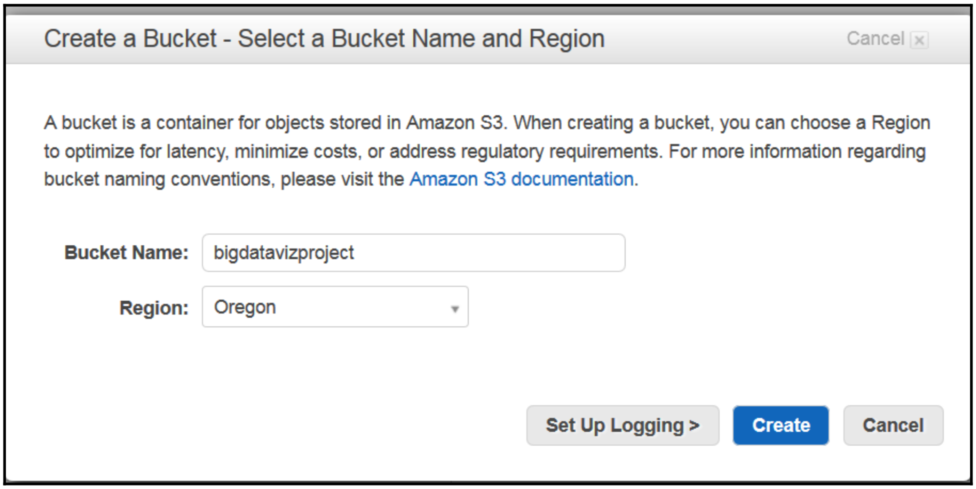
Welcome to Amazon Simple Storage Service

Amazon S3 is storage for the Internet. It is designed to make web-scale computing easier for developers.

Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, secure, fast, inexpensive infrastructure that Amazon uses to run its own global network of web sites. The service aims to maximize benefits of scale and to pass those benefits on to developers.

You can read, write, and delete objects ranging in size from 1 byte to 5 terabytes each. The number of objects you can store is unlimited. Each object is stored in a bucket with a unique key that you assign.

Get started by simply creating a bucket and uploading a test object, for example a photo or .txt file.

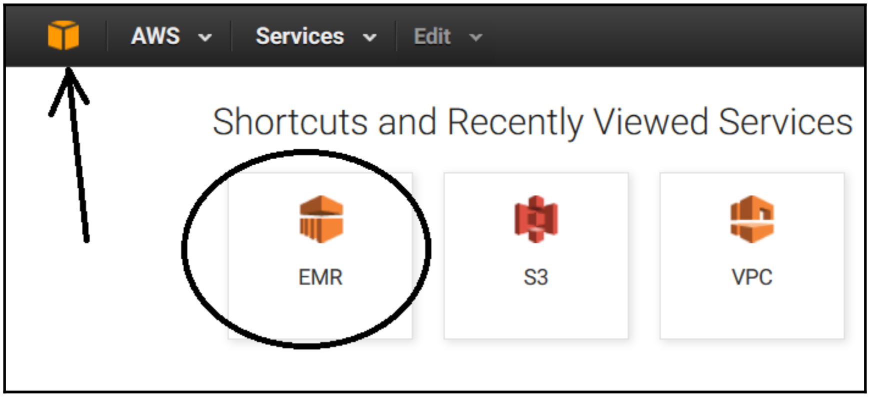


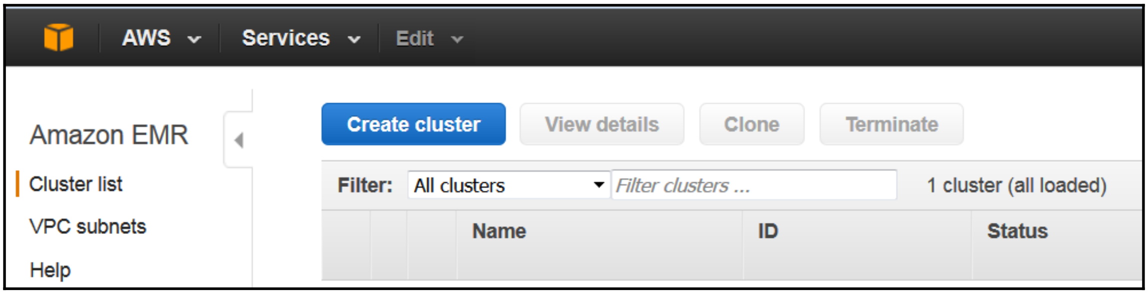
Create a Bucket - Select a Bucket Name and Region

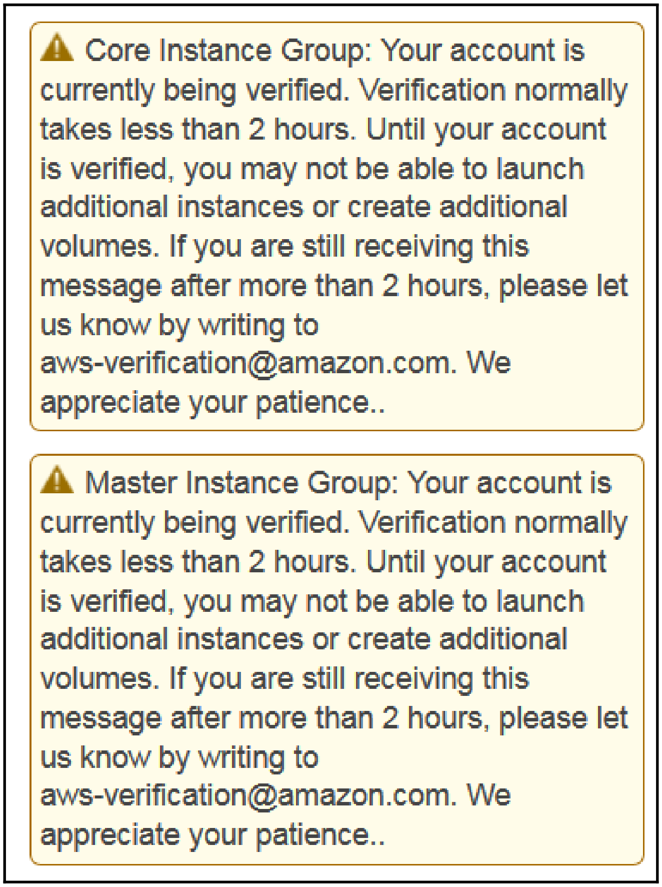
A bucket is a container for objects stored in Amazon S3. When creating a bucket, you can choose a Region to optimize for latency, minimize costs, or address regulatory requirements. For more information regarding bucket naming conventions, please visit the Amazon S3 documentation.

Bucket Name: bigdatavizproject

Region: Oregon

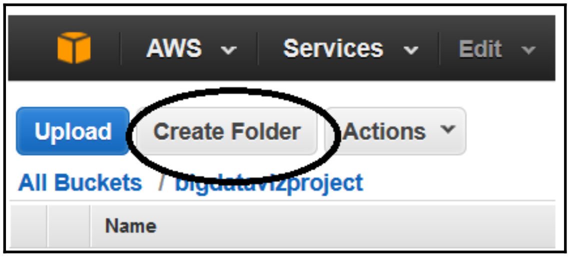

AWS
Services
Edit
Shortcuts and Recently Viewed Services
EMR
S3
VPC

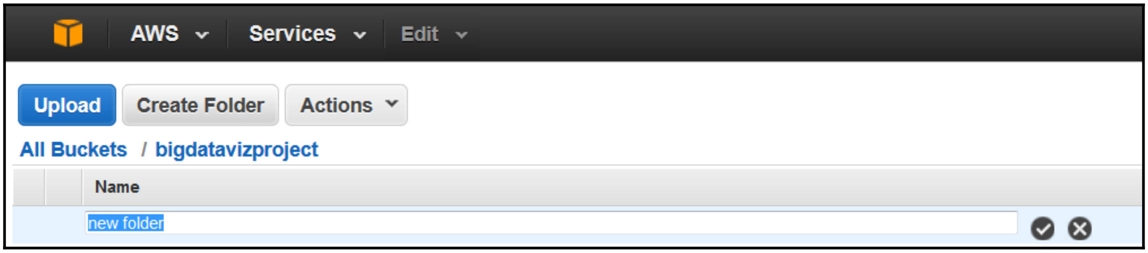

AWS
Services
Edit
Amazon EMR
Cluster list
VPC subnets
Help
Create cluster
View details
Clone
Terminate
Filter:
All clusters
Filter clusters ...
1 cluster (all loaded)
Name
ID
Status



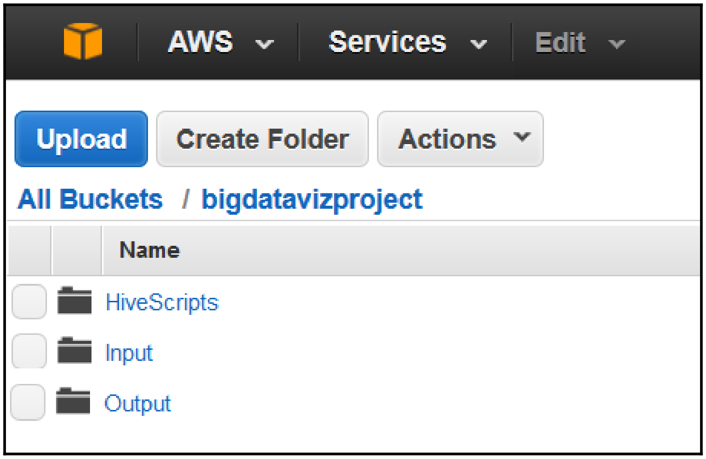
⚠ Core Instance Group: Your account is currently being verified. Verification normally takes less than 2 hours. Until your account is verified, you may not be able to launch additional instances or create additional volumes. If you are still receiving this message after more than 2 hours, please let us know by writing to aws-verification@amazon.com. We appreciate your patience..

⚠ Master Instance Group: Your account is currently being verified. Verification normally takes less than 2 hours. Until your account is verified, you may not be able to launch additional instances or create additional volumes. If you are still receiving this message after more than 2 hours, please let us know by writing to aws-verification@amazon.com. We appreciate your patience..

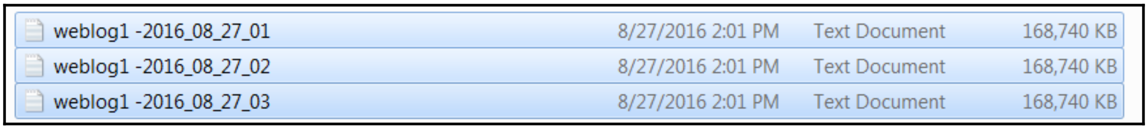




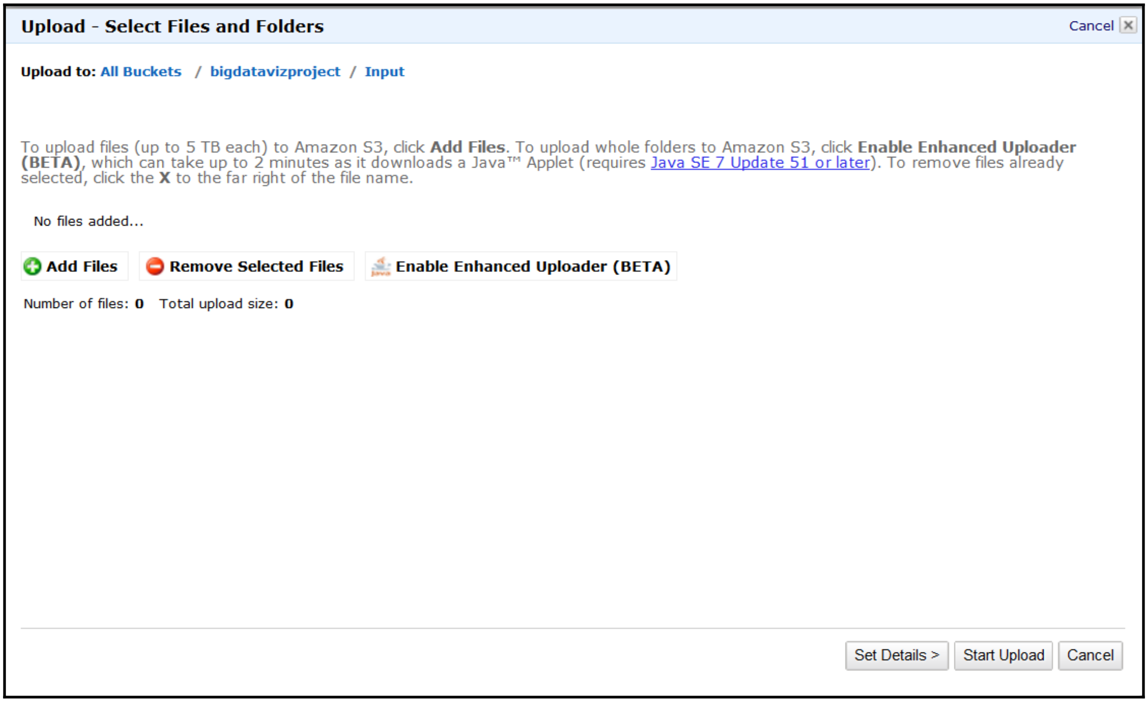
new folder



HiveScripts
Input
Output



weblog1 -2016_08_27_01	8/27/2016 2:01 PM	Text Document	168,740 KB
weblog1 -2016_08_27_02	8/27/2016 2:01 PM	Text Document	168,740 KB
weblog1 -2016_08_27_03	8/27/2016 2:01 PM	Text Document	168,740 KB



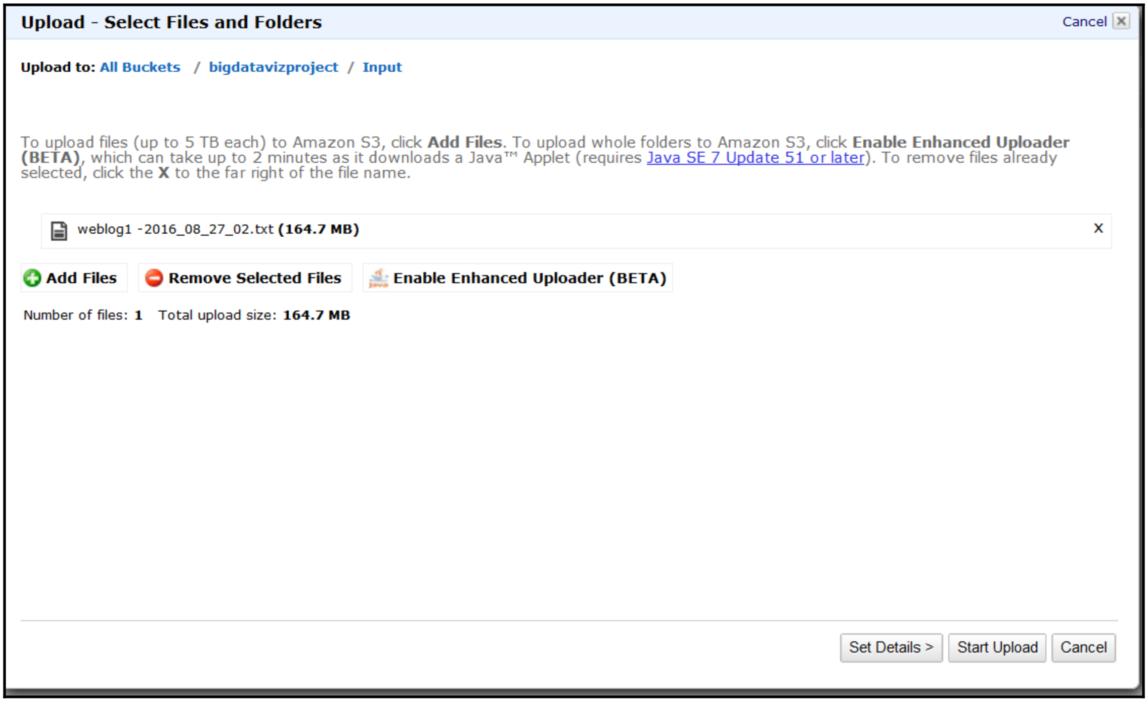
Upload - Select Files and Folders

Upload to: All Buckets / bigdatavizproject / Input

To upload files (up to 5 TB each) to Amazon S3, click **Add Files**. To upload whole folders to Amazon S3, click **Enable Enhanced Uploader (BETA)**, which can take up to 2 minutes as it downloads a Java™ Applet (requires Java SE 7 Update 51 or later). To remove files already selected, click the **X** to the far right of the file name.

No files added...

Number of files: **0** Total upload size: **0**



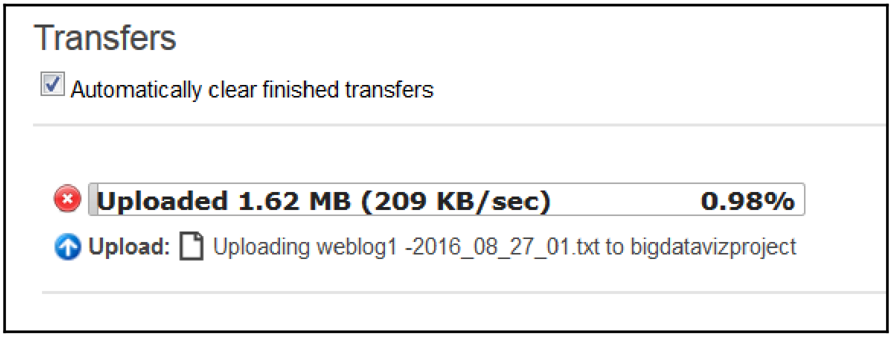
Upload - Select Files and Folders

Upload to: All Buckets / bigdatavizproject / Input

To upload files (up to 5 TB each) to Amazon S3, click **Add Files**. To upload whole folders to Amazon S3, click **Enable Enhanced Uploader (BETA)**, which can take up to 2 minutes as it downloads a Java™ Applet (requires Java SE 7 Update 51 or later). To remove files already selected, click the **X** to the far right of the file name.

weblog1 -2016_08_27_02.txt **(164.7 MB)** X

Number of files: **1** Total upload size: **164.7 MB**

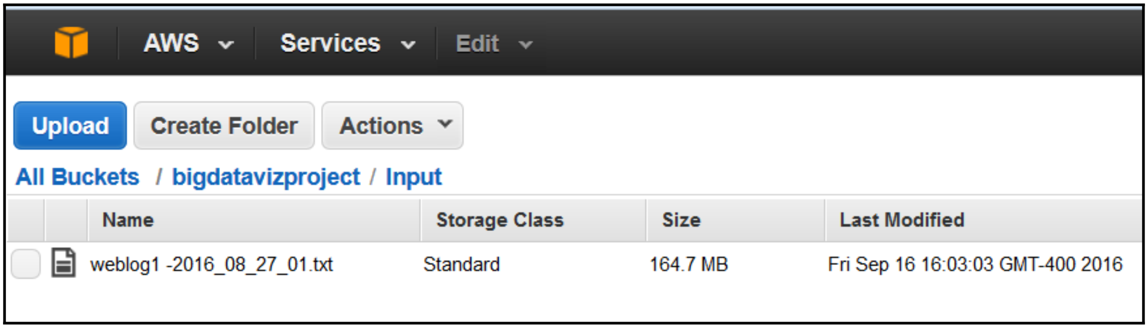


Transfers

☑ Automatically clear finished transfers

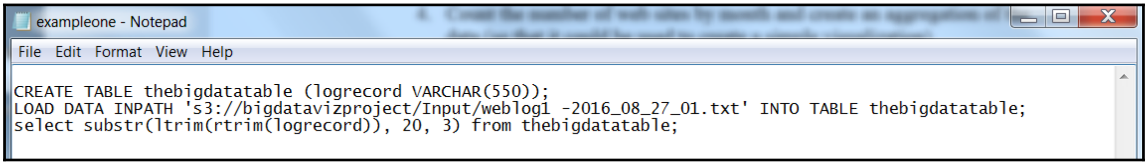
Uploaded 1.62 MB (209 KB/sec) 0.98%

Upload: Uploading weblog1 -2016_08_27_01.txt to bigdatavizproject

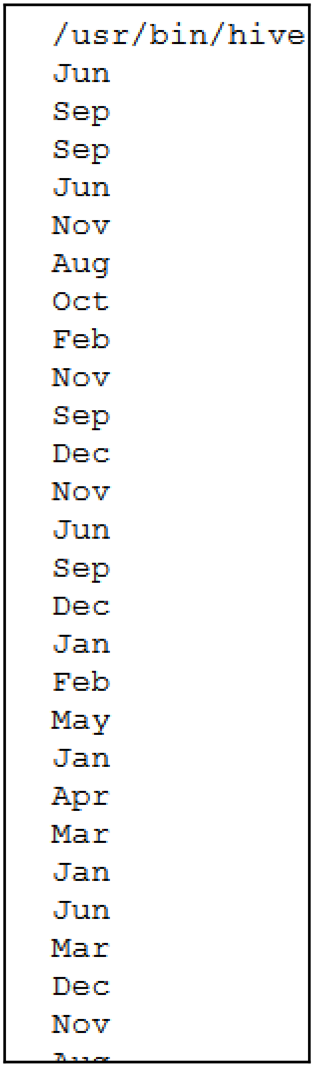


All Buckets / bigdatavizproject / Input

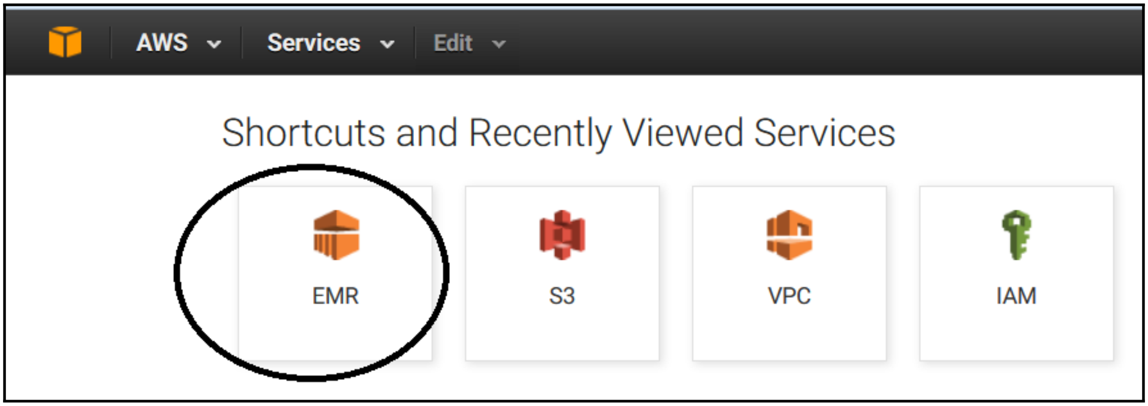
Name	Storage Class	Size	Last Modified
weblog1 -2016_08_27_01.txt	Standard	164.7 MB	Fri Sep 16 16:03:03 GMT-400 2016

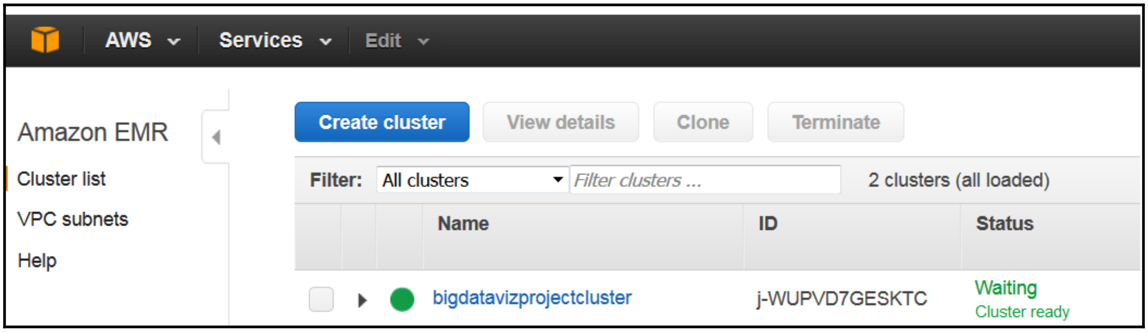


```
CREATE TABLE thebigdatatable (logrecord VARCHAR(550));
LOAD DATA INPATH 's3://bigdatavizproject/Input/weblog1 -2016_08_27_01.txt' INTO TABLE thebigdatatable;
select substr(ltrim(rtrim(logrecord)), 20, 3) from thebigdatatable;
```



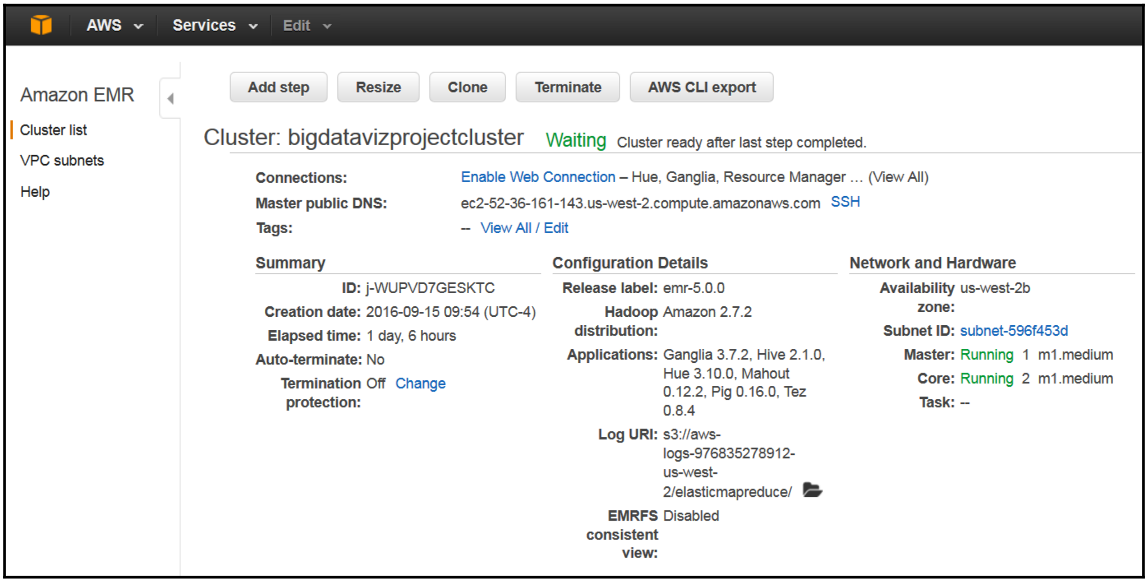
```
/usr/bin/hive
Jun
Sep
Sep
Jun
Nov
Aug
Oct
Feb
Nov
Sep
Dec
Nov
Jun
Sep
Dec
Jan
Feb
May
Jan
Apr
Mar
Jan
Jun
Mar
Dec
Nov
Aug
```





Filter: All clusters | Filter clusters ... | 2 clusters (all loaded)

			Name	ID	Status
□	▸	●	bigdatavizprojectcluster	j-WUPVD7GESKTC	Waiting Cluster ready



Cluster: bigdatavizprojectcluster Waiting Cluster ready after last step completed.

Connections: Enable Web Connection – Hue, Ganglia, Resource Manager ... (View All)
Master public DNS: ec2-52-36-161-143.us-west-2.compute.amazonaws.com SSH
Tags: -- View All / Edit

Summary

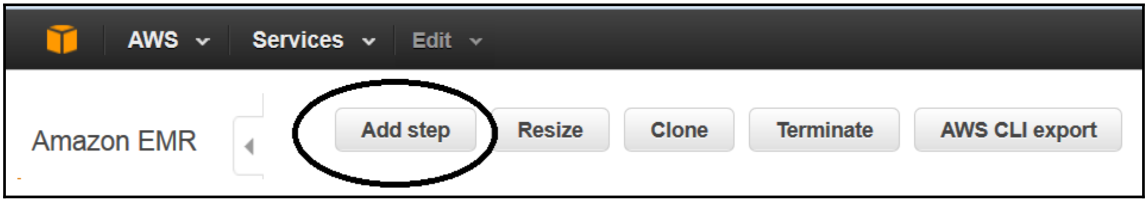
ID: j-WUPVD7GESKTC
Creation date: 2016-09-15 09:54 (UTC-4)
Elapsed time: 1 day, 6 hours
Auto-terminate: No
Termination protection: Off Change

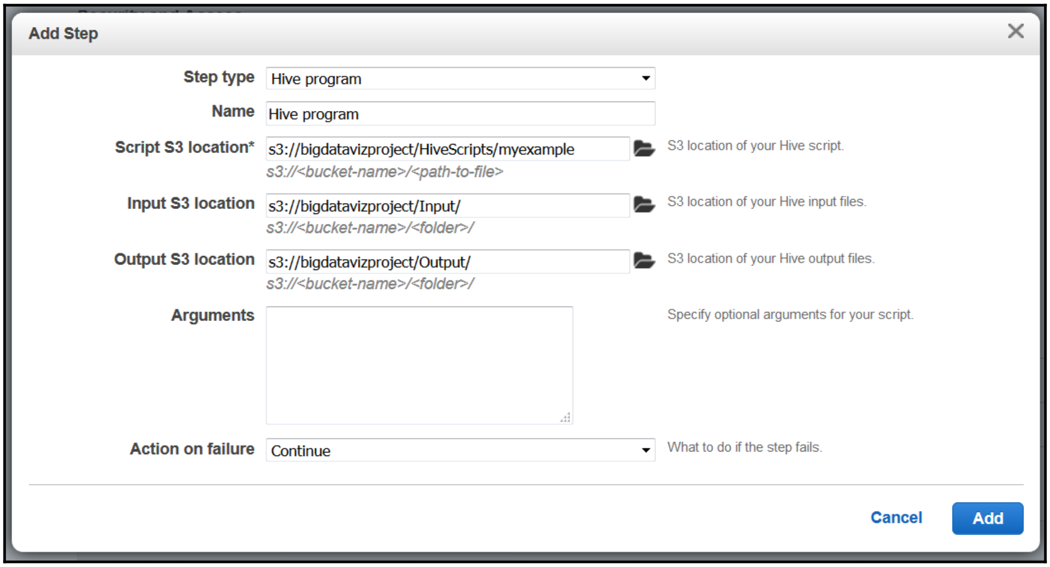
Configuration Details

Release label: emr-5.0.0
Hadoop distribution: Amazon 2.7.2
Applications: Ganglia 3.7.2, Hive 2.1.0, Hue 3.10.0, Mahout 0.12.2, Pig 0.16.0, Tez 0.8.4
Log URI: s3://aws-logs-976835278912-us-west-2/elasticmapreduce/
EMRFS consistent view: Disabled

Network and Hardware

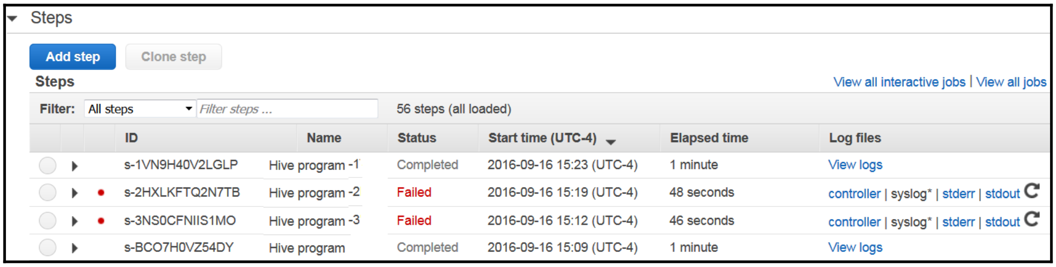
Availability zone: us-west-2b
Subnet ID: subnet-596f453d
Master: Running 1 m1.medium
Core: Running 2 m1.medium
Task: --





Add Step

Step type	Hive program	
Name	Hive program	
Script S3 location*	s3://bigdatavizproject/HiveScripts/myexample *s3://<bucket-name>/<path-to-file>*	S3 location of your Hive script.
Input S3 location	s3://bigdatavizproject/Input/ *s3://<bucket-name>/<folder>/*	S3 location of your Hive input files.
Output S3 location	s3://bigdatavizproject/Output/ *s3://<bucket-name>/<folder>/*	S3 location of your Hive output files.
Arguments		Specify optional arguments for your script.
Action on failure	Continue	What to do if the step fails.

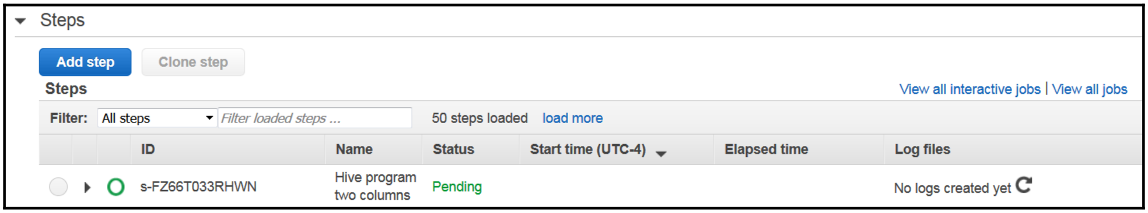


Steps

View all interactive jobs | View all jobs

Filter: All steps *Filter steps ...* 56 steps (all loaded)

ID	Name	Status	Start time (UTC-4)	Elapsed time	Log files
s-1VN9H40V2LGLP	Hive program -1	Completed	2016-09-16 15:23 (UTC-4)	1 minute	View logs
● s-2HXLKFTQ2N7TB	Hive program -2	Failed	2016-09-16 15:19 (UTC-4)	48 seconds	controller | syslog* | stderr | stdout
● s-3NS0CFNIIS1MO	Hive program -3	Failed	2016-09-16 15:12 (UTC-4)	46 seconds	controller | syslog* | stderr | stdout
s-BCO7H0VZ54DY	Hive program	Completed	2016-09-16 15:09 (UTC-4)	1 minute	View logs

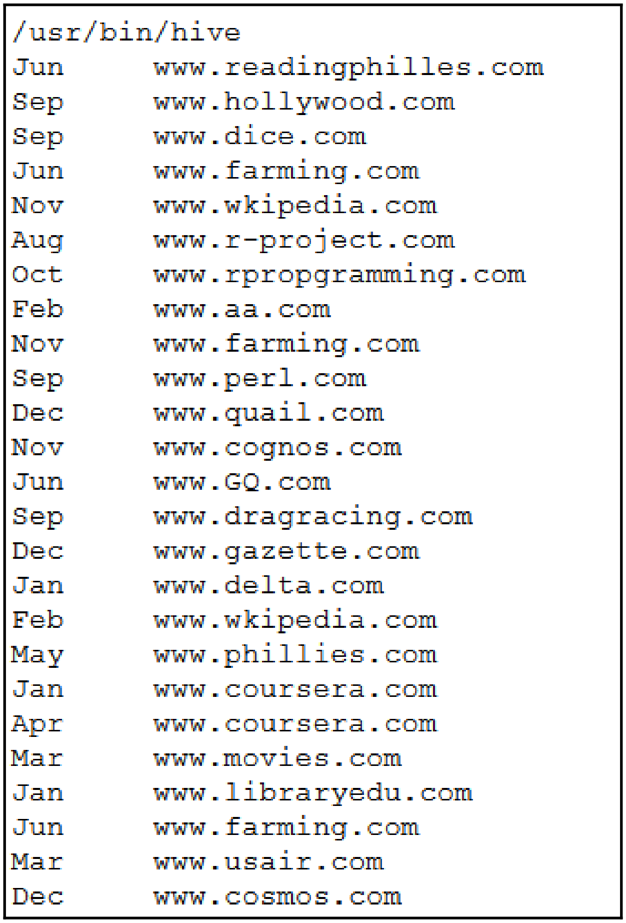


Steps

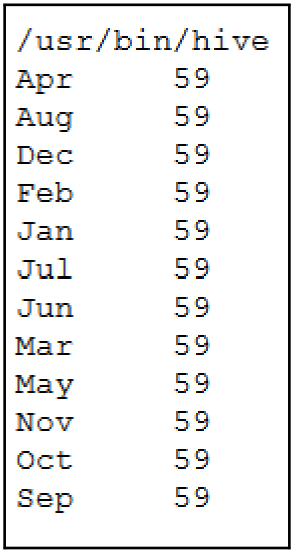
View all interactive jobs | View all jobs

Filter: All steps *Filter loaded steps ...* 50 steps loaded load more

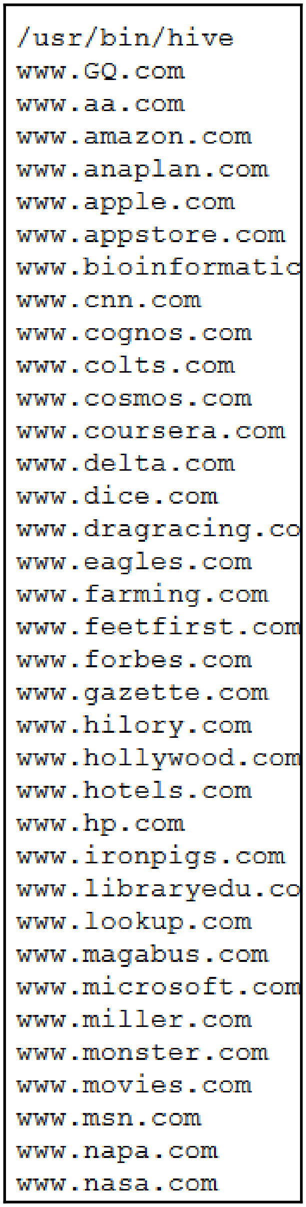
ID	Name	Status	Start time (UTC-4)	Elapsed time	Log files
○ s-FZ66T033RHWN	Hive program two columns	Pending			No logs created yet



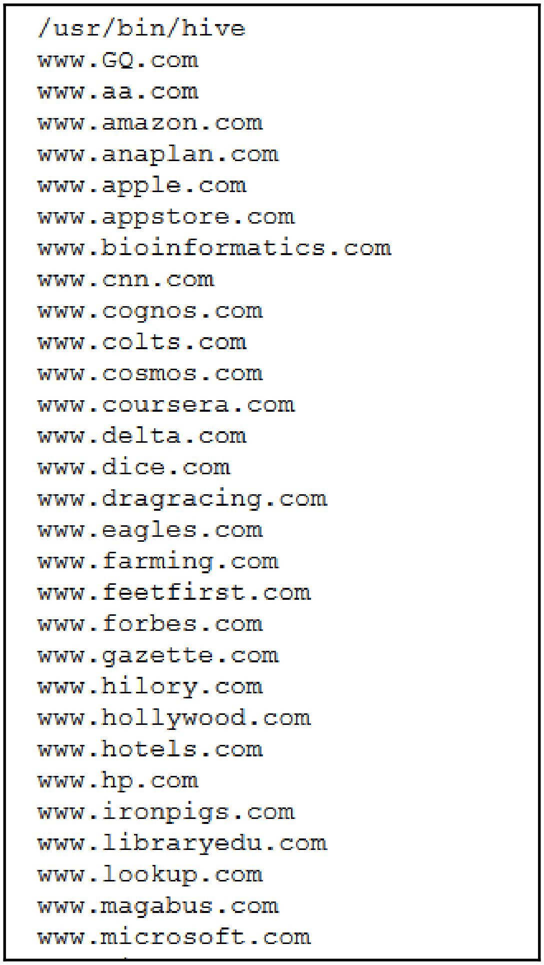
```
/usr/bin/hive
Jun     www.readingphilles.com
Sep     www.hollywood.com
Sep     www.dice.com
Jun     www.farming.com
Nov     www.wkipedia.com
Aug     www.r-project.com
Oct     www.rpropgramming.com
Feb     www.aa.com
Nov     www.farming.com
Sep     www.perl.com
Dec     www.quail.com
Nov     www.cognos.com
Jun     www.GQ.com
Sep     www.dragracing.com
Dec     www.gazette.com
Jan     www.delta.com
Feb     www.wkipedia.com
May     www.phillies.com
Jan     www.coursera.com
Apr     www.coursera.com
Mar     www.movies.com
Jan     www.libraryedu.com
Jun     www.farming.com
Mar     www.usair.com
Dec     www.cosmos.com
```



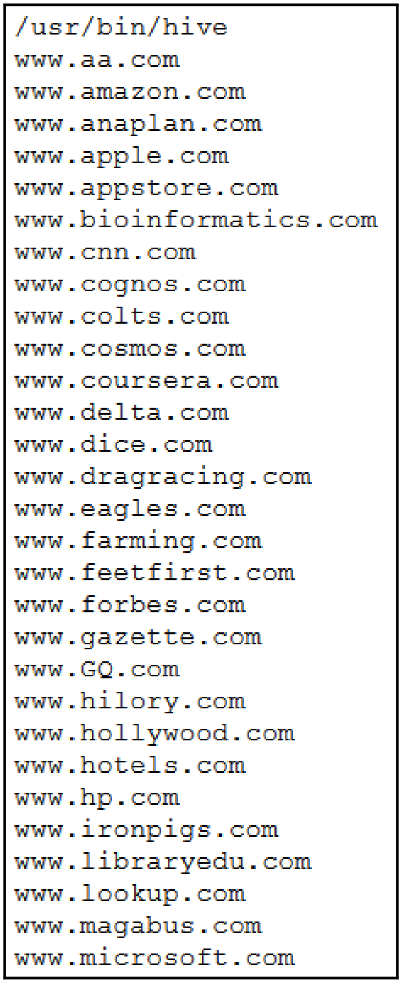
```
/usr/bin/hive
Apr     59
Aug     59
Dec     59
Feb     59
Jan     59
Jul     59
Jun     59
Mar     59
May     59
Nov     59
Oct     59
Sep     59
```



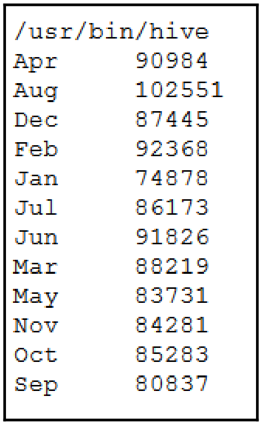
```
/usr/bin/hive
www.GQ.com
www.aa.com
www.amazon.com
www.anaplan.com
www.apple.com
www.appstore.com
www.bioinformatic
www.cnn.com
www.cognos.com
www.colts.com
www.cosmos.com
www.coursera.com
www.delta.com
www.dice.com
www.dragracing.co
www.eagles.com
www.farming.com
www.feetfirst.com
www.forbes.com
www.gazette.com
www.hilory.com
www.hollywood.com
www.hotels.com
www.hp.com
www.ironpigs.com
www.libraryedu.co
www.lookup.com
www.magabus.com
www.microsoft.com
www.miller.com
www.monster.com
www.movies.com
www.msn.com
www.napa.com
www.nasa.com
```



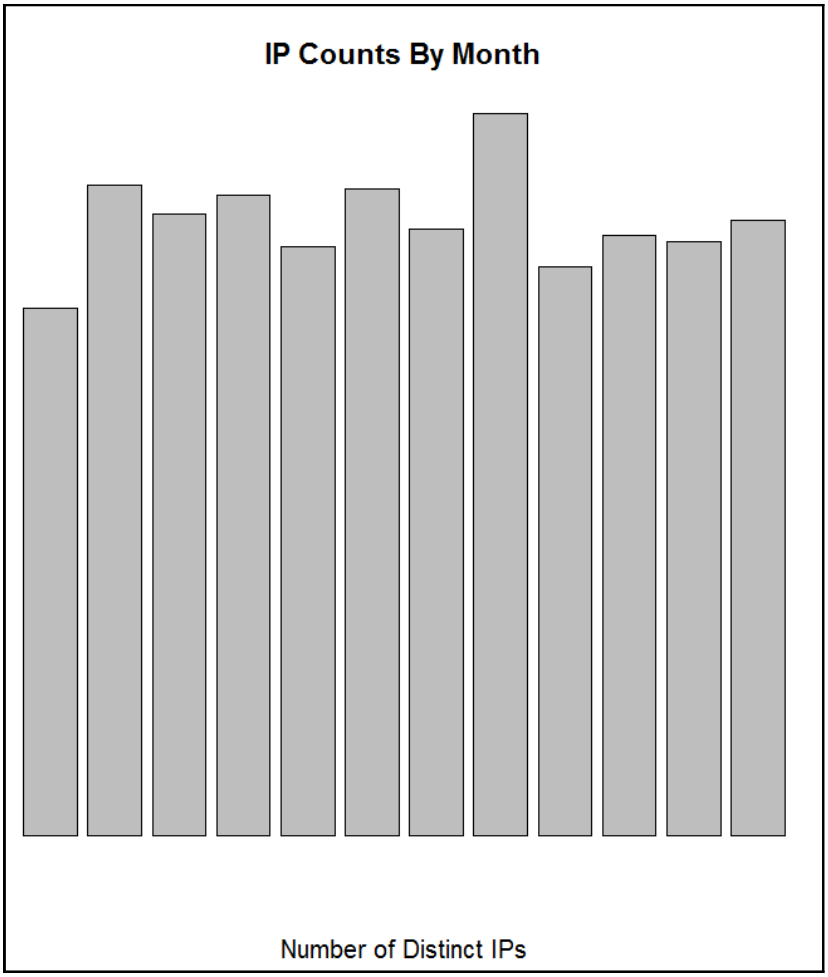
```
/usr/bin/hive
www.GQ.com
www.aa.com
www.amazon.com
www.anaplan.com
www.apple.com
www.appstore.com
www.bioinformatics.com
www.cnn.com
www.cognos.com
www.colts.com
www.cosmos.com
www.coursera.com
www.delta.com
www.dice.com
www.dragracing.com
www.eagles.com
www.farming.com
www.feetfirst.com
www.forbes.com
www.gazette.com
www.hilory.com
www.hollywood.com
www.hotels.com
www.hp.com
www.ironpigs.com
www.libraryedu.com
www.lookup.com
www.magabus.com
www.microsoft.com
```



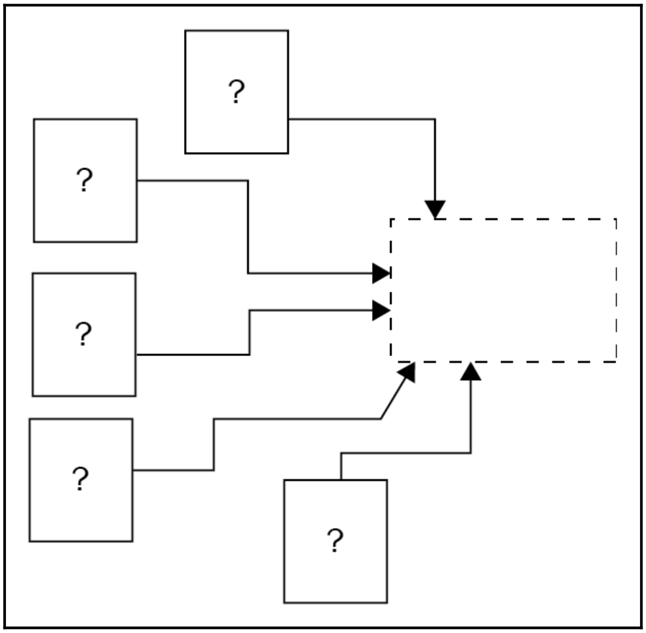
```
/usr/bin/hive
www.aa.com
www.amazon.com
www.anaplan.com
www.apple.com
www.appstore.com
www.bioinformatics.com
www.cnn.com
www.cognos.com
www.colts.com
www.cosmos.com
www.coursera.com
www.delta.com
www.dice.com
www.dragracing.com
www.eagles.com
www.farming.com
www.feetfirst.com
www.forbes.com
www.gazette.com
www.GQ.com
www.hilory.com
www.hollywood.com
www.hotels.com
www.hp.com
www.ironpigs.com
www.libraryedu.com
www.lookup.com
www.magabus.com
www.microsoft.com
```

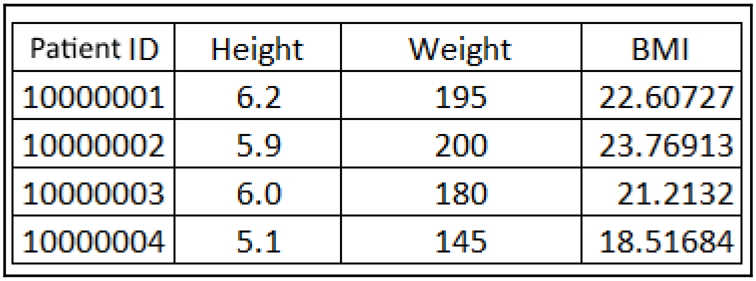


```
/usr/bin/hive
Apr       90984
Aug       102551
Dec       87445
Feb       92368
Jan       74878
Jul       86173
Jun       91826
Mar       88219
May       83731
Nov       84281
Oct       85283
Sep       80837
```

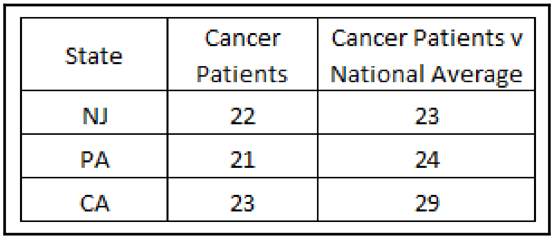

IP Counts By Month
Number of Distinct IPs

Chapter 3: Understanding Your Data Using R

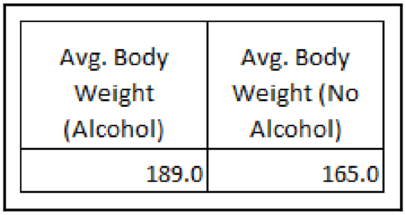




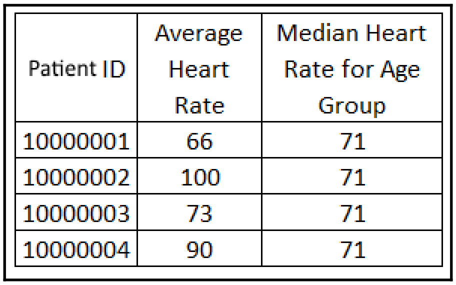
Patient ID	Height	Weight	BMI
10000001	6.2	195	22.60727
10000002	5.9	200	23.76913
10000003	6.0	180	21.2132
10000004	5.1	145	18.51684



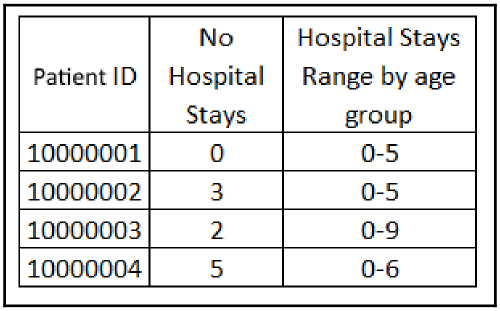
State	Cancer Patients	Cancer Patients v National Average
NJ	22	23
PA	21	24
CA	23	29



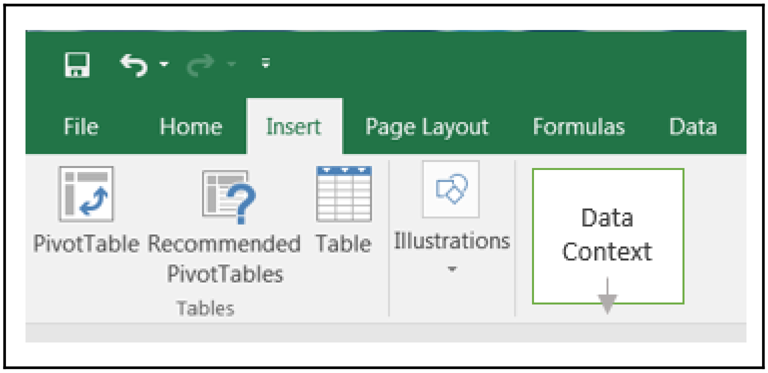
Avg. Body Weight (Alcohol)	Avg. Body Weight (No Alcohol)
189.0	165.0

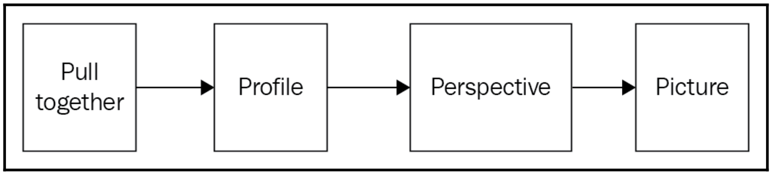


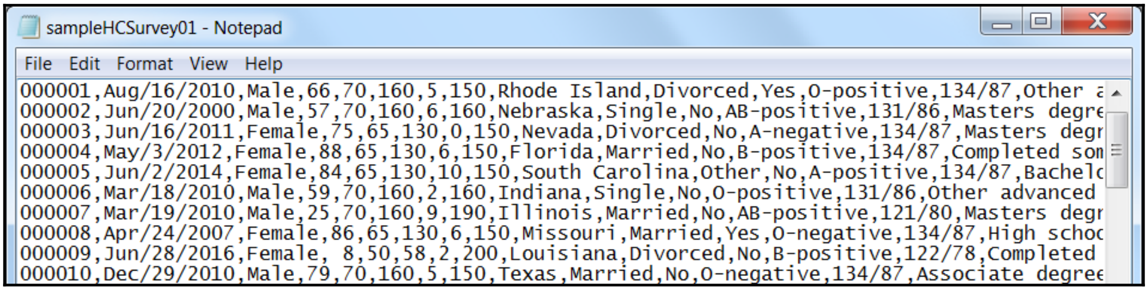
Patient ID	Average Heart Rate	Median Heart Rate for Age Group
10000001	66	71
10000002	100	71
10000003	73	71
10000004	90	71



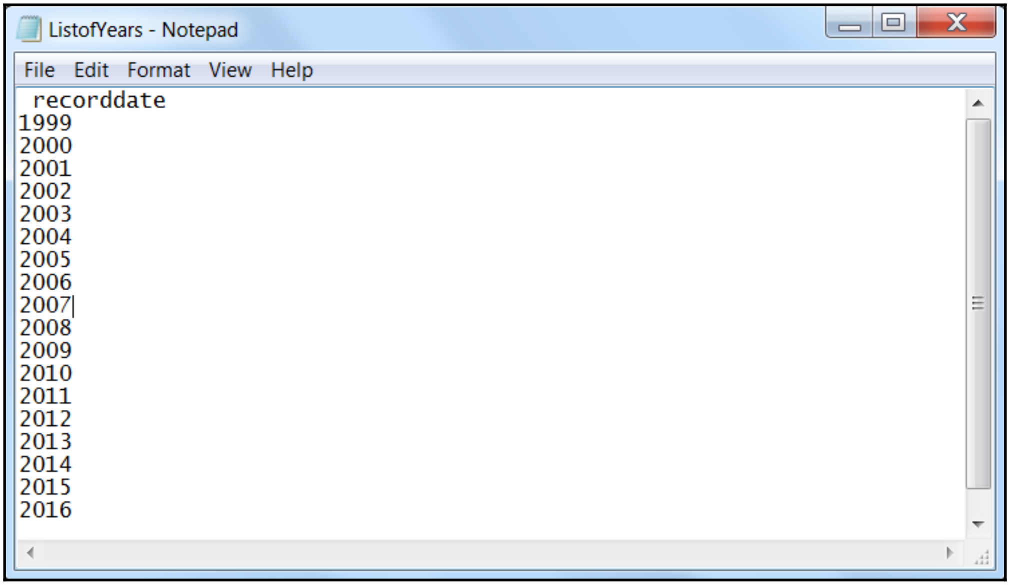
Patient ID	No Hospital Stays	Hospital Stays Range by age group
10000001	0	0-5
10000002	3	0-5
10000003	2	0-9
10000004	5	0-6



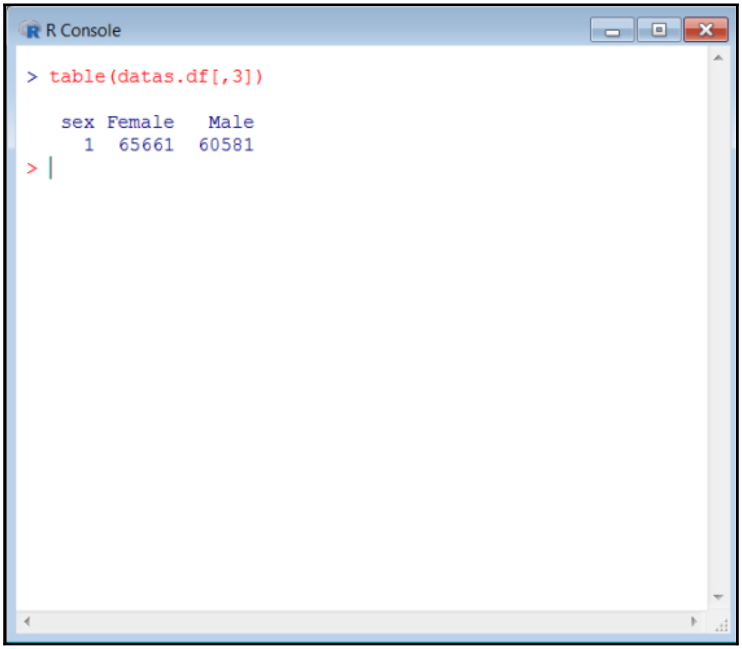




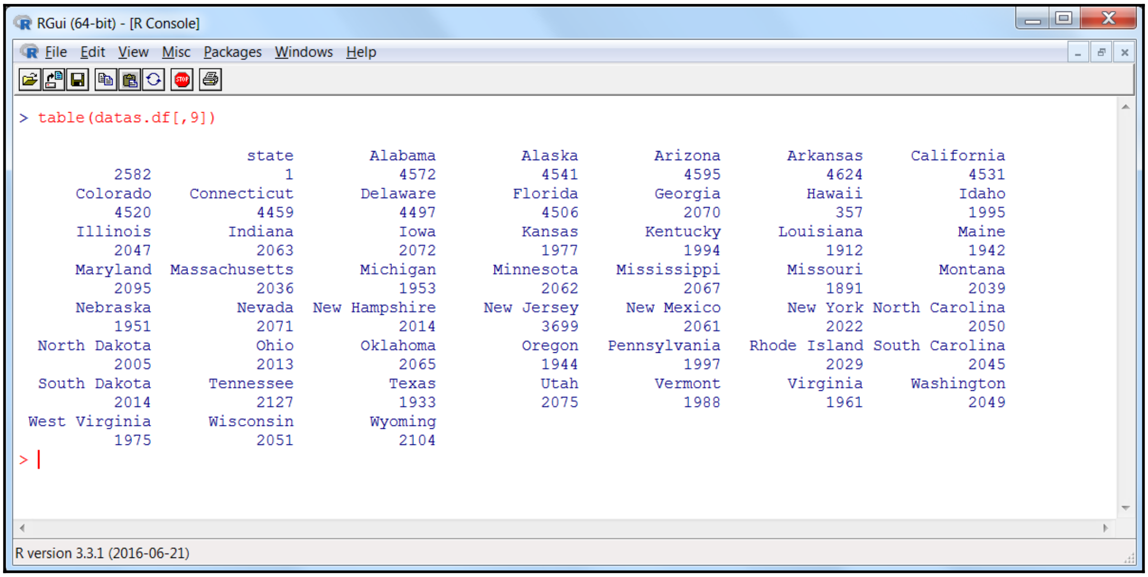
```
000001,Aug/16/2010,Male,66,70,160,5,150,Rhode Island,Divorced,Yes,O-positive,134/87,Other a
000002,Jun/20/2000,Male,57,70,160,6,160,Nebraska,Single,No,AB-positive,131/86,Masters degre
000003,Jun/16/2011,Female,75,65,130,0,150,Nevada,Divorced,No,A-negative,134/87,Masters degr
000004,May/3/2012,Female,88,65,130,6,150,Florida,Married,No,B-positive,134/87,Completed som
000005,Jun/2/2014,Female,84,65,130,10,150,South Carolina,Other,No,A-positive,134/87,Bachelo
000006,Mar/18/2010,Male,59,70,160,2,160,Indiana,Single,No,O-positive,131/86,Other advanced
000007,Mar/19/2010,Male,25,70,160,9,190,Illinois,Married,No,AB-positive,121/80,Masters degr
000008,Apr/24/2007,Female,86,65,130,6,150,Missouri,Married,Yes,O-negative,134/87,High scho
000009,Jun/28/2016,Female, 8,50,58,2,200,Louisiana,Divorced,No,B-positive,122/78,Completed
000010,Dec/29/2010,Male,79,70,160,5,150,Texas,Married,No,O-negative,134/87,Associate degree
```



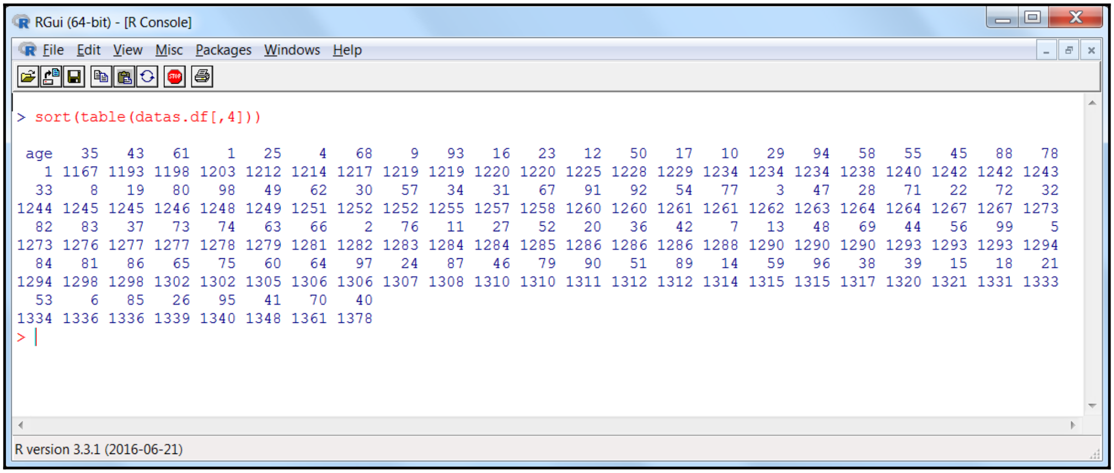
```
 recorddate
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
```



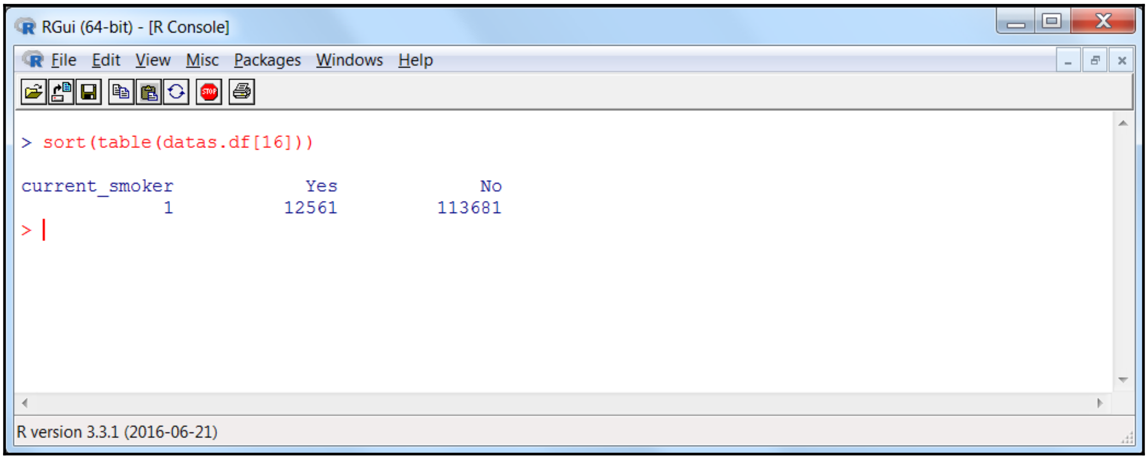
```
> table(datas.df[,3])

   sex Female   Male
     1  65661  60581
> 
```


RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
R version 3.3.1 (2016-06-21)

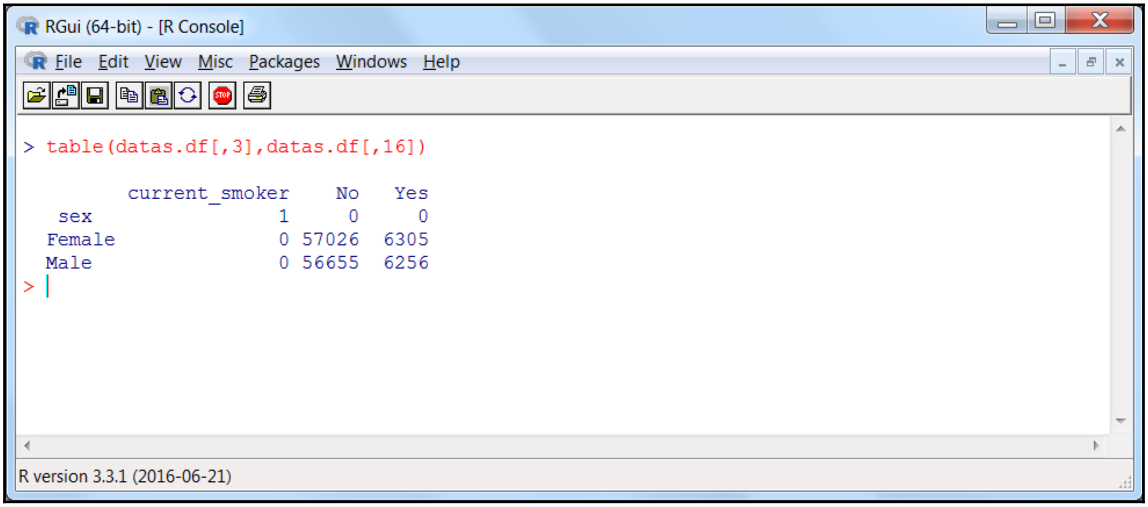
```
> table(datas.df[,9])

                         state        Alabama         Alaska        Arizona       Arkansas     California
          2582               1           4572           4541           4595           4624           4531
      Colorado     Connecticut       Delaware        Florida        Georgia         Hawaii          Idaho
          4520            4459           4497           4506           2070            357           1995
      Illinois         Indiana           Iowa         Kansas       Kentucky      Louisiana          Maine
          2047            2063           2072           1977           1994           1912           1942
      Maryland   Massachusetts       Michigan      Minnesota    Mississippi       Missouri        Montana
          2095            2036           1953           2062           2067           1891           2039
      Nebraska          Nevada  New Hampshire     New Jersey     New Mexico       New York North Carolina
          1951            2071           2014           3699           2061           2022           2050
  North Dakota            Ohio       Oklahoma         Oregon   Pennsylvania   Rhode Island South Carolina
          2005            2013           2065           1944           1997           2029           2045
  South Dakota       Tennessee          Texas           Utah        Vermont       Virginia     Washington
          2014            2127           1933           2075           1988           1961           2049
 West Virginia       Wisconsin        Wyoming
          1975            2051           2104
> 
```


RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
R version 3.3.1 (2016-06-21)

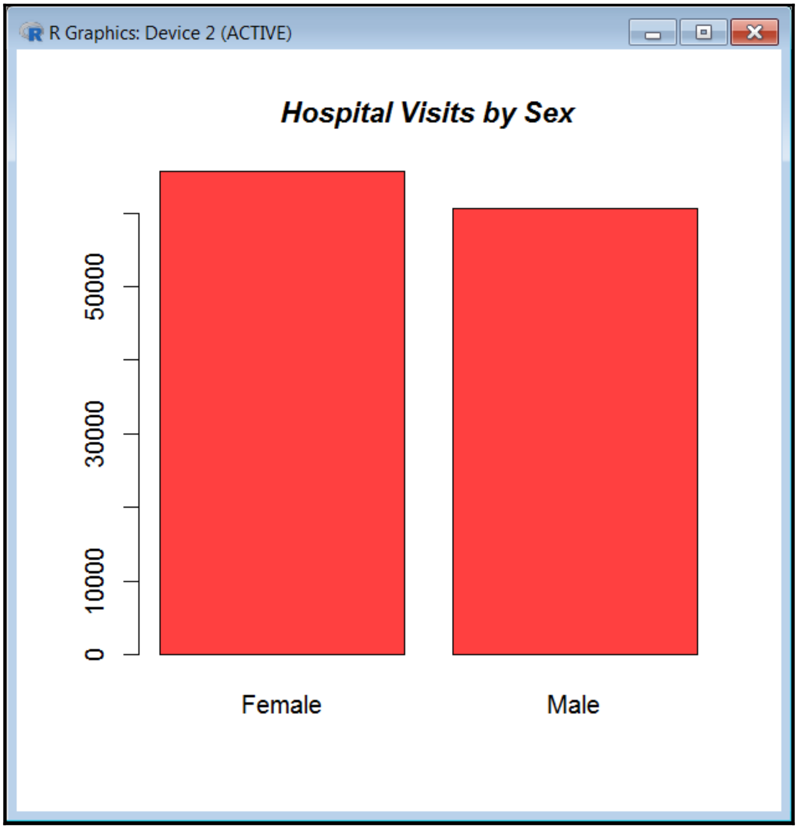
```
> sort(table(datas.df[,4]))

 age   35   43   61    1   25    4   68    9   93   16   23   12   50   17   10   29   94   58   55   45   88   78 
   1 1167 1193 1198 1203 1212 1214 1217 1219 1219 1220 1220 1225 1228 1229 1234 1234 1234 1238 1240 1242 1242 1243 
  33    8   19   80   98   49   62   30   57   34   31   67   91   92   54   77    3   47   28   71   22   72   32 
1244 1245 1245 1246 1248 1249 1251 1252 1252 1255 1257 1258 1260 1260 1261 1261 1262 1263 1264 1264 1267 1267 1273 
  82   83   37   73   74   63   66    2   76   11   27   52   20   36   42    7   13   48   69   44   56   99    5 
1273 1276 1277 1277 1278 1279 1281 1282 1283 1284 1284 1285 1286 1286 1286 1288 1290 1290 1290 1293 1293 1293 1294 
  84   81   86   65   75   60   64   97   24   87   46   79   90   51   89   14   59   96   38   39   15   18   21 
1294 1298 1298 1302 1302 1305 1306 1306 1307 1308 1310 1310 1311 1312 1312 1314 1315 1315 1317 1320 1321 1331 1333 
  53    6   85   26   95   41   70   40 
1334 1336 1336 1339 1340 1348 1361 1378 
> 
```

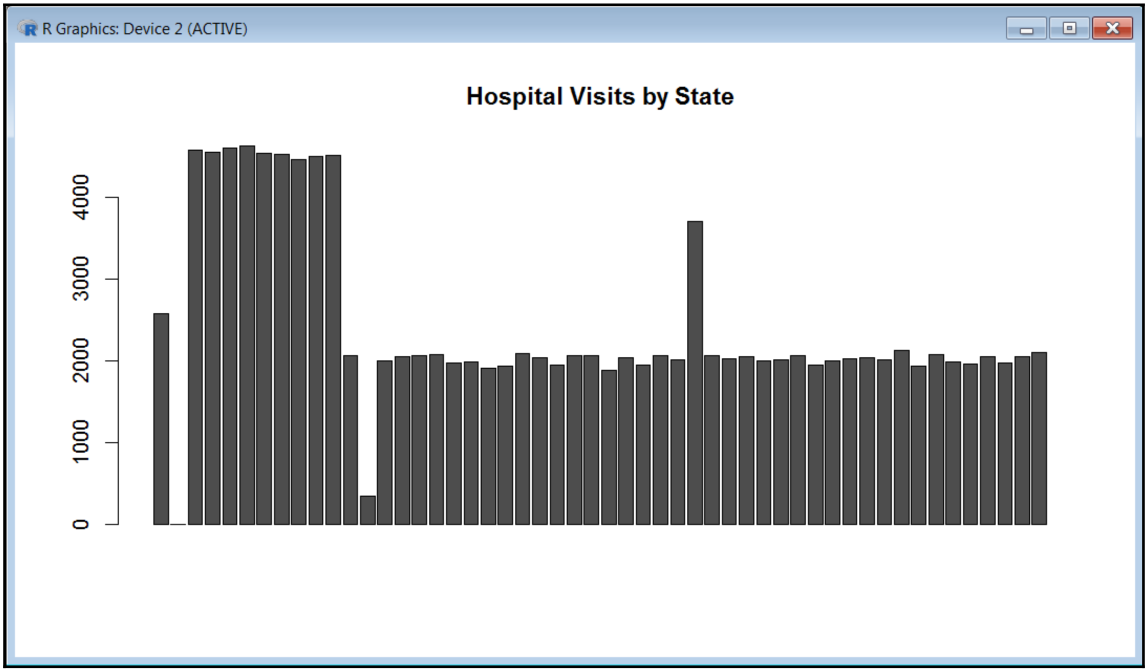

RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
R version 3.3.1 (2016-06-21)

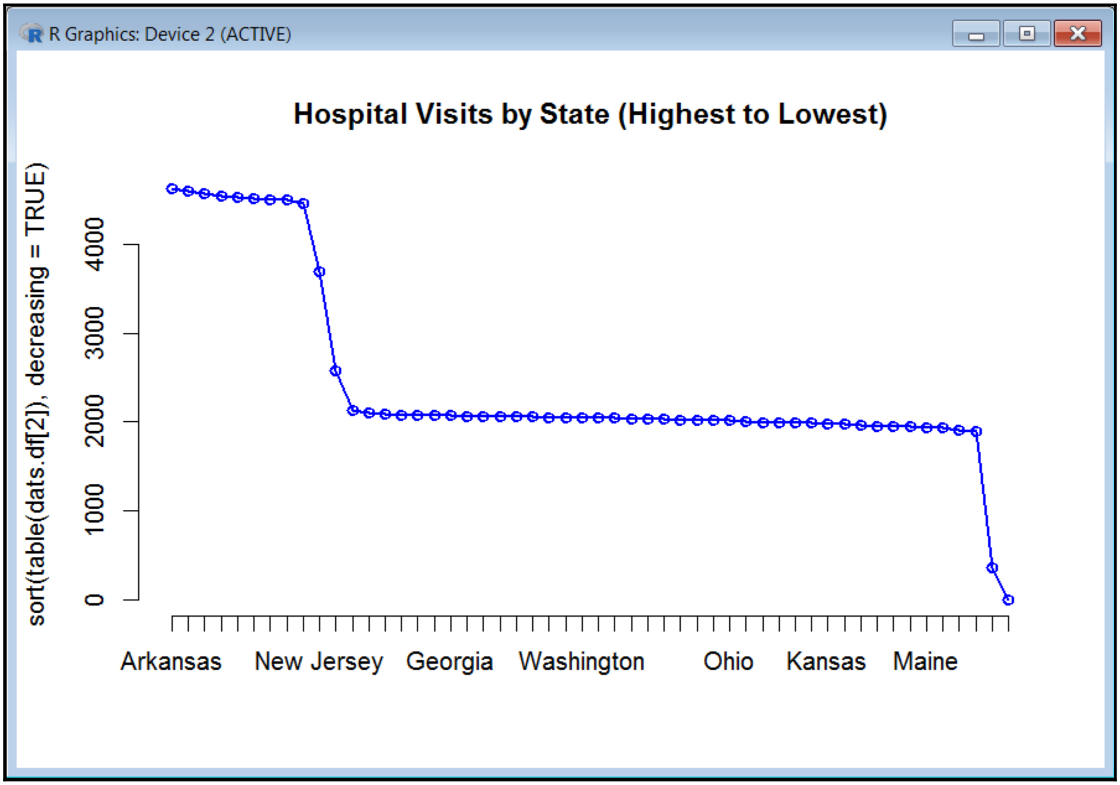
```
> sort(table(datas.df[16]))

current_smoker            Yes             No 
             1          12561         113681 
> 
```


RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
R version 3.3.1 (2016-06-21)

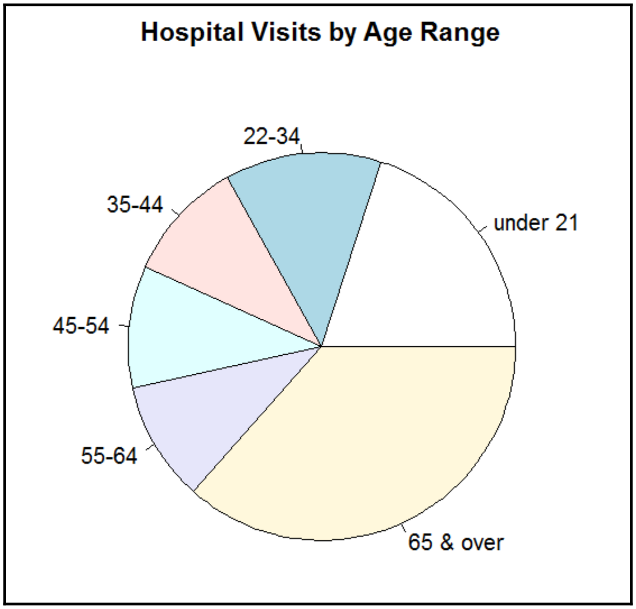
```
> table(datas.df[,3],datas.df[,16])

        current_smoker    No   Yes
  sex                1     0     0
 Female              0 57026  6305
 Male                0 56655  6256
> 
```

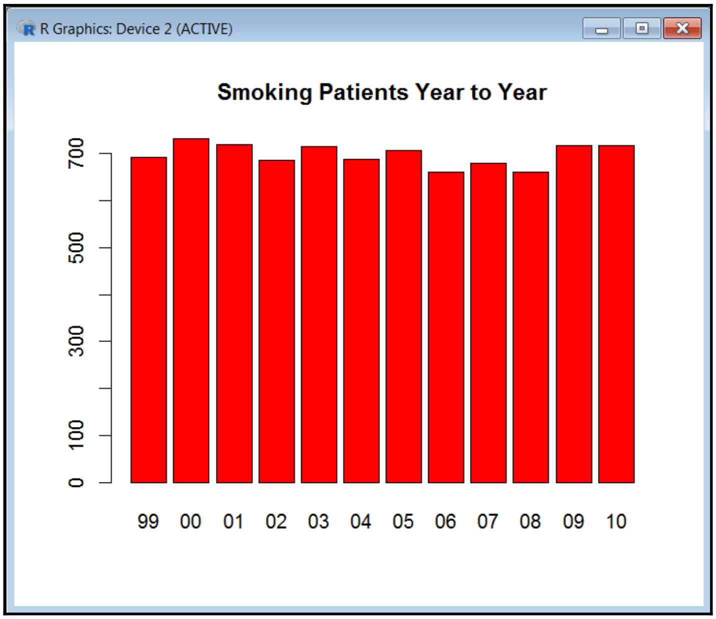


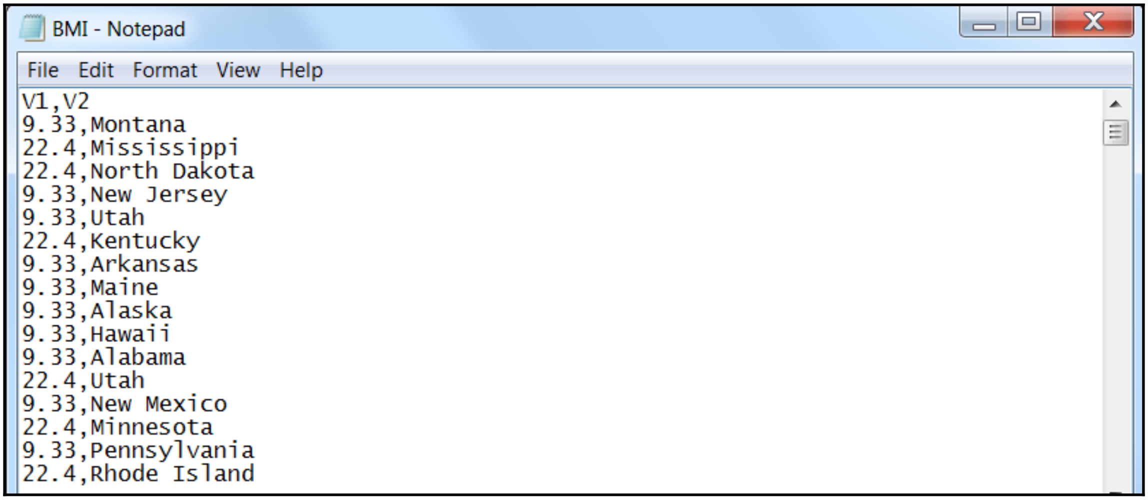

R Graphics: Device 2 (ACTIVE)
Hospital Visits by State
0
1000
2000
3000
4000



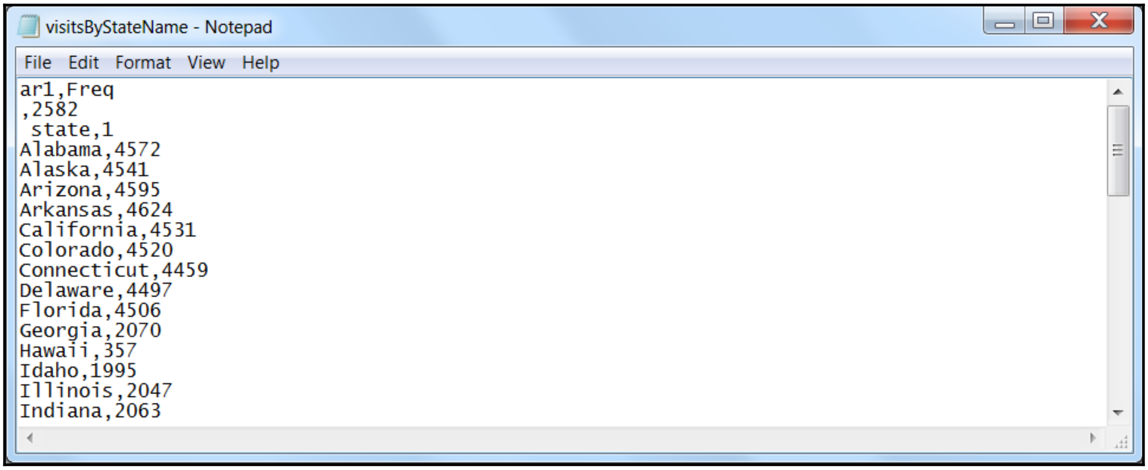
R Graphics: Device 2 (ACTIVE)
Hospital Visits by State (Highest to Lowest)
sort(table(dats.df[2]), decreasing = TRUE)
0
1000
2000
3000
4000
Arkansas
New Jersey
Georgia
Washington
Ohio
Kansas
Maine


Hospital Visits by Age Range
22-34
under 21
35-44
45-54
55-64
65 & over

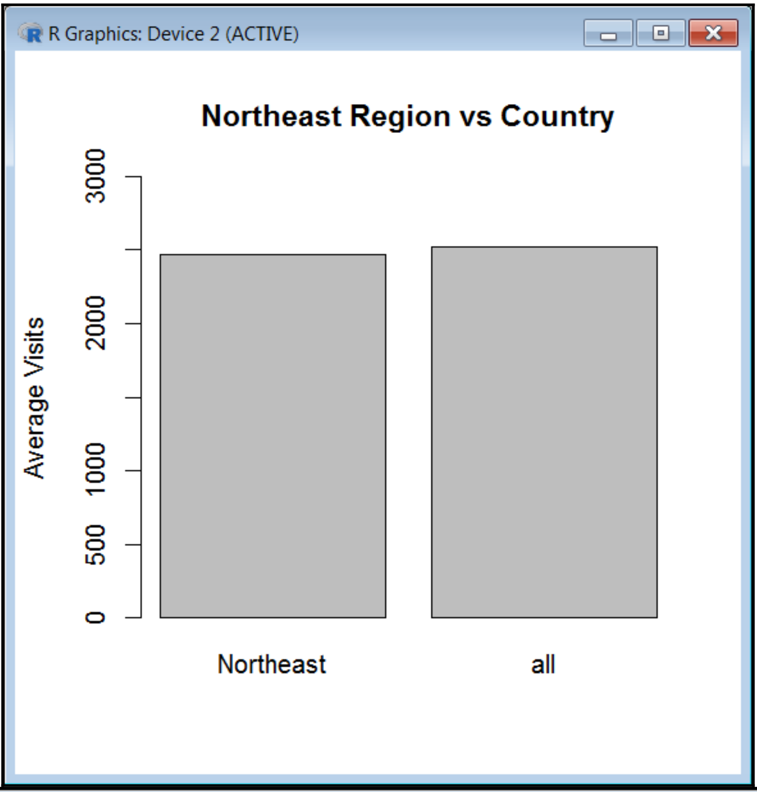

R Graphics: Device 2 (ACTIVE)
Smoking Patients Year to Year
0
100
300
500
700
99
00
01
02
03
04
05
06
07
08
09
10

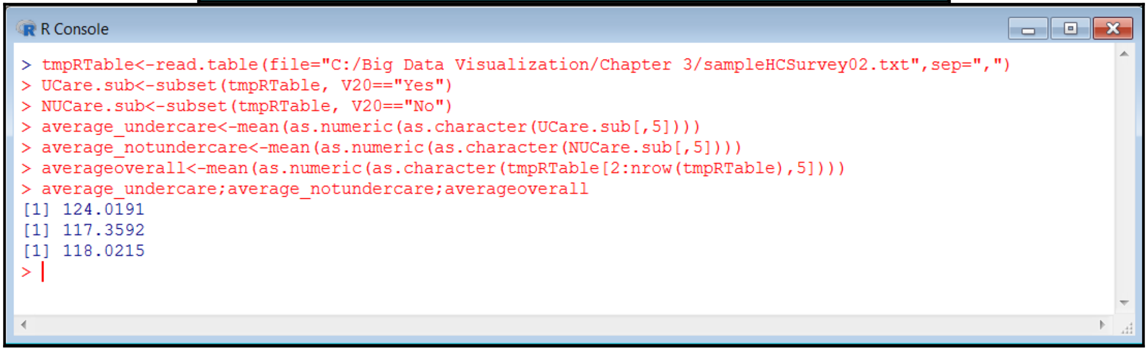


```
V1,V2
9.33,Montana
22.4,Mississippi
22.4,North Dakota
9.33,New Jersey
9.33,Utah
22.4,Kentucky
9.33,Arkansas
9.33,Maine
9.33,Alaska
9.33,Hawaii
9.33,Alabama
22.4,Utah
9.33,New Mexico
22.4,Minnesota
9.33,Pennsylvania
22.4,Rhode Island
```

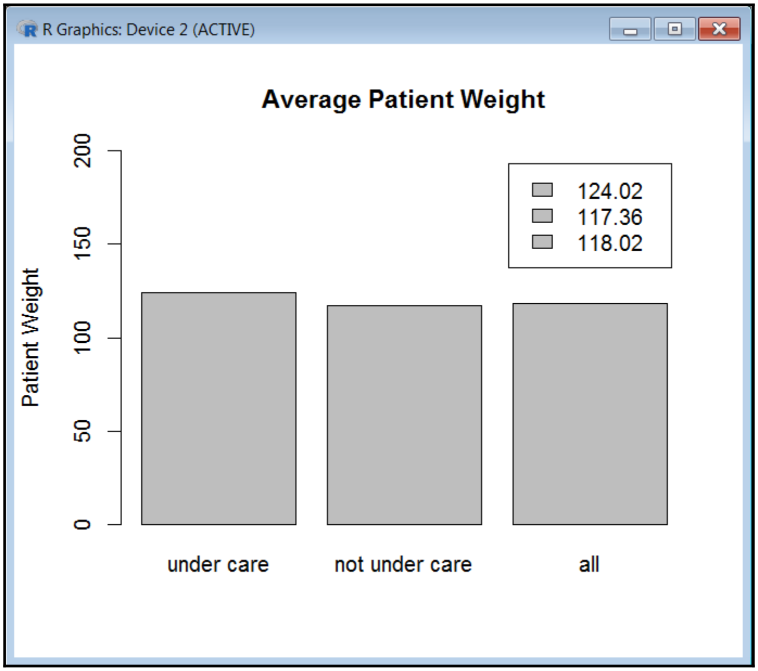


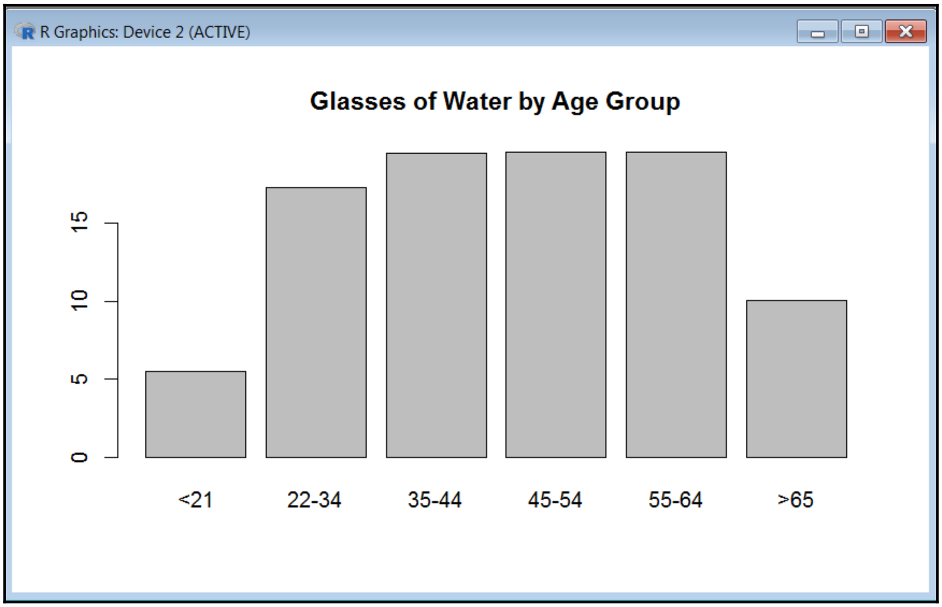
```
ar1,Freq
,2582
 state,1
Alabama,4572
Alaska,4541
Arizona,4595
Arkansas,4624
California,4531
Colorado,4520
Connecticut,4459
Delaware,4497
Florida,4506
Georgia,2070
Hawaii,357
Idaho,1995
Illinois,2047
Indiana,2063
```

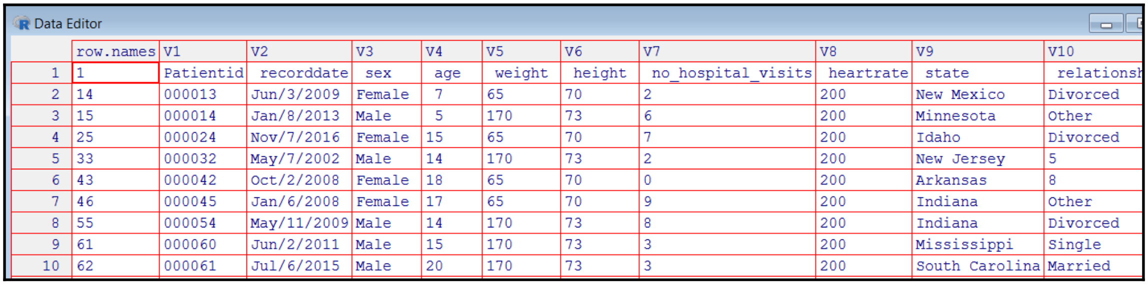




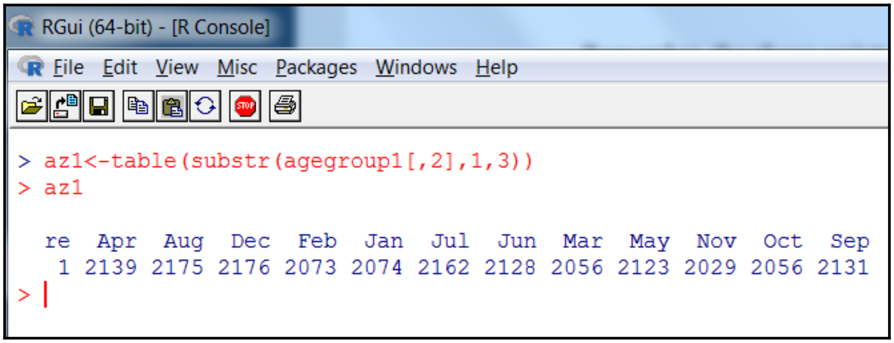
```
> tmpRTable<-read.table(file="C:/Big Data Visualization/Chapter 3/sampleHCSurvey02.txt",sep=",")
> UCare.sub<-subset(tmpRTable, V20=="Yes")
> NUCare.sub<-subset(tmpRTable, V20=="No")
> average_undercare<-mean(as.numeric(as.character(UCare.sub[,5])))
> average_notundercare<-mean(as.numeric(as.character(NUCare.sub[,5])))
> averageoverall<-mean(as.numeric(as.character(tmpRTable[2:nrow(tmpRTable),5])))
> average_undercare;average_notundercare;averageoverall
[1] 124.0191
[1] 117.3592
[1] 118.0215
> 
```


R Graphics: Device 2 (ACTIVE)
Average Patient Weight
Patient Weight
0
50
100
150
200
🟫 124.02
🟫 117.36
🟫 118.02
under care
not under care
all

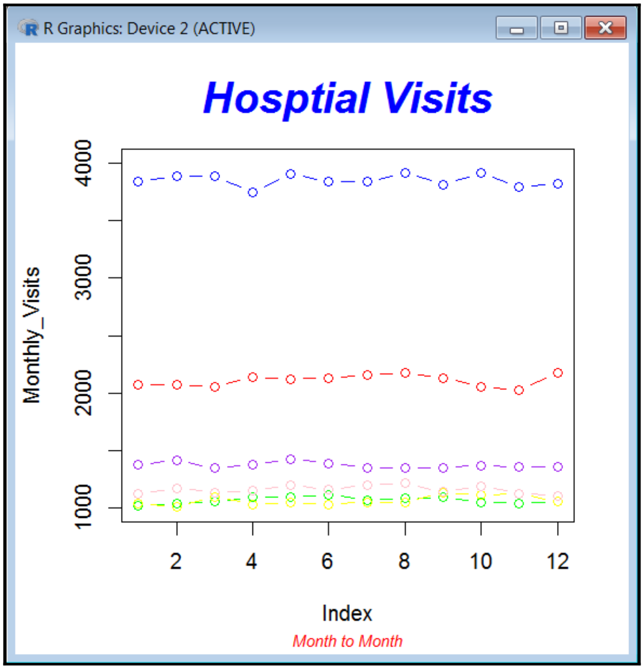

R Graphics: Device 2 (ACTIVE)
Glasses of Water by Age Group
0
5
10
15
<21
22-34
35-44
45-54
55-64
>65



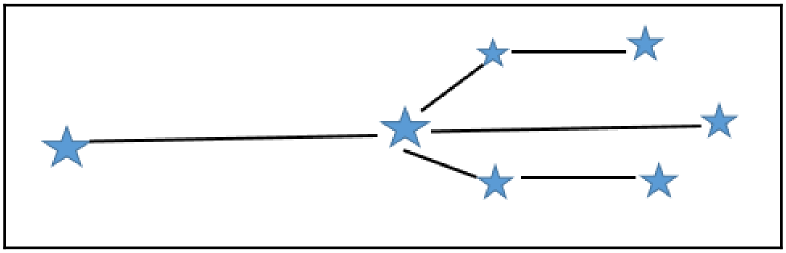
	row.names	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	1	Patientid	recorddate	sex	age	weight	height	no_hospital_visits	heartrate	state	relationsh
2	14	000013	Jun/3/2009	Female	7	65	70	2	200	New Mexico	Divorced
3	15	000014	Jan/8/2013	Male	5	170	73	6	200	Minnesota	Other
4	25	000024	Nov/7/2016	Female	15	65	70	7	200	Idaho	Divorced
5	33	000032	May/7/2002	Male	14	170	73	2	200	New Jersey	5
6	43	000042	Oct/2/2008	Female	18	65	70	0	200	Arkansas	8
7	46	000045	Jan/6/2008	Female	17	65	70	9	200	Indiana	Other
8	55	000054	May/11/2009	Male	14	170	73	8	200	Indiana	Divorced
9	61	000060	Jun/2/2011	Male	15	170	73	3	200	Mississippi	Single
10	62	000061	Jul/6/2015	Male	20	170	73	3	200	South Carolina	Married

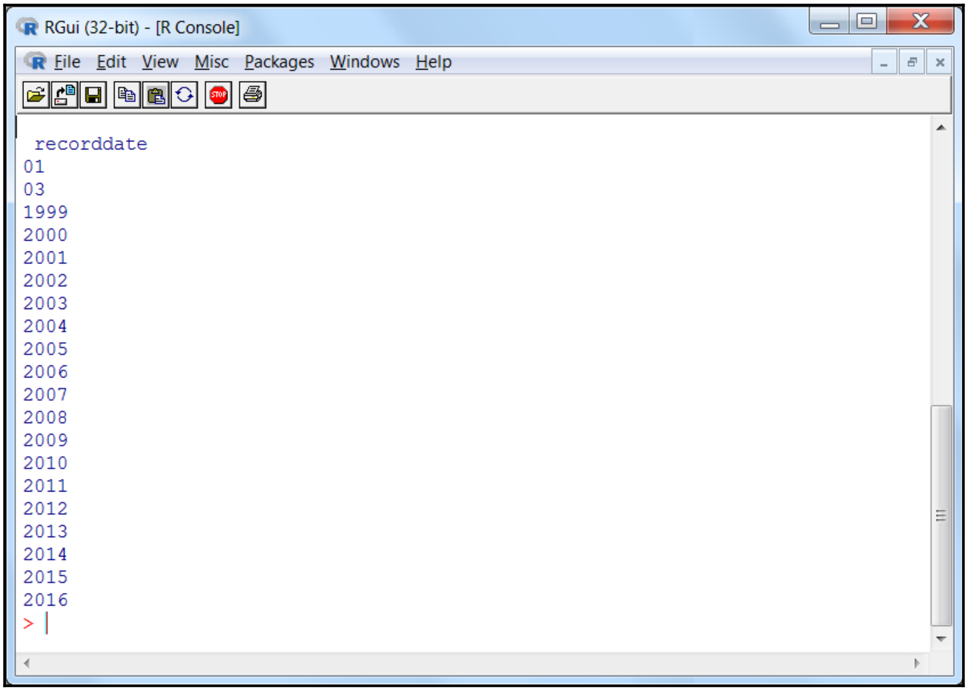

RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

```
> az1<-table(substr(agegroup1[,2],1,3))
> az1

  re  Apr  Aug  Dec  Feb  Jan  Jul  Jun  Mar  May  Nov  Oct  Sep 
   1 2139 2175 2176 2073 2074 2162 2128 2056 2123 2029 2056 2131 
> 
```

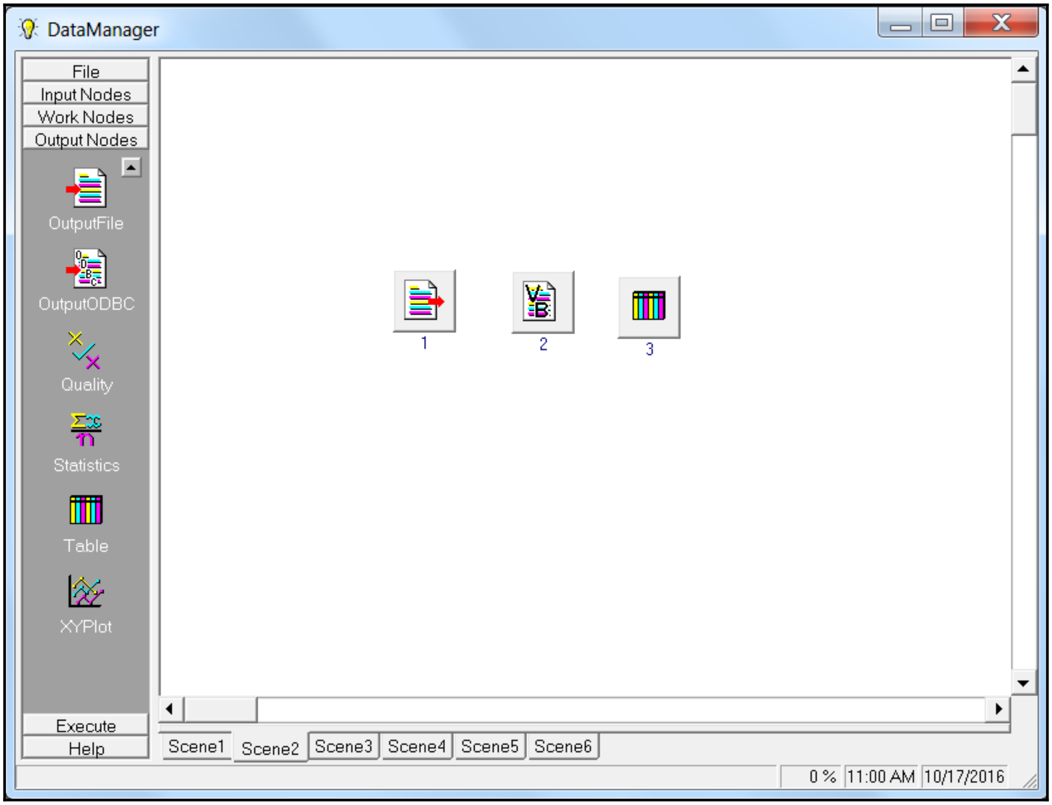


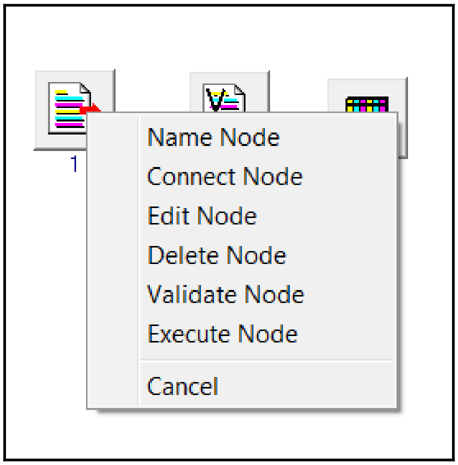
Chapter 4: Addressing Big Data Quality

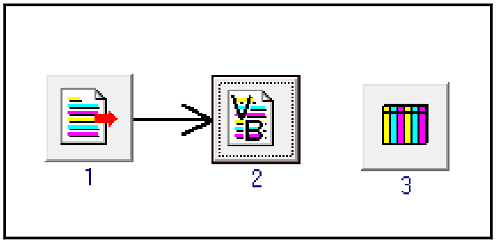


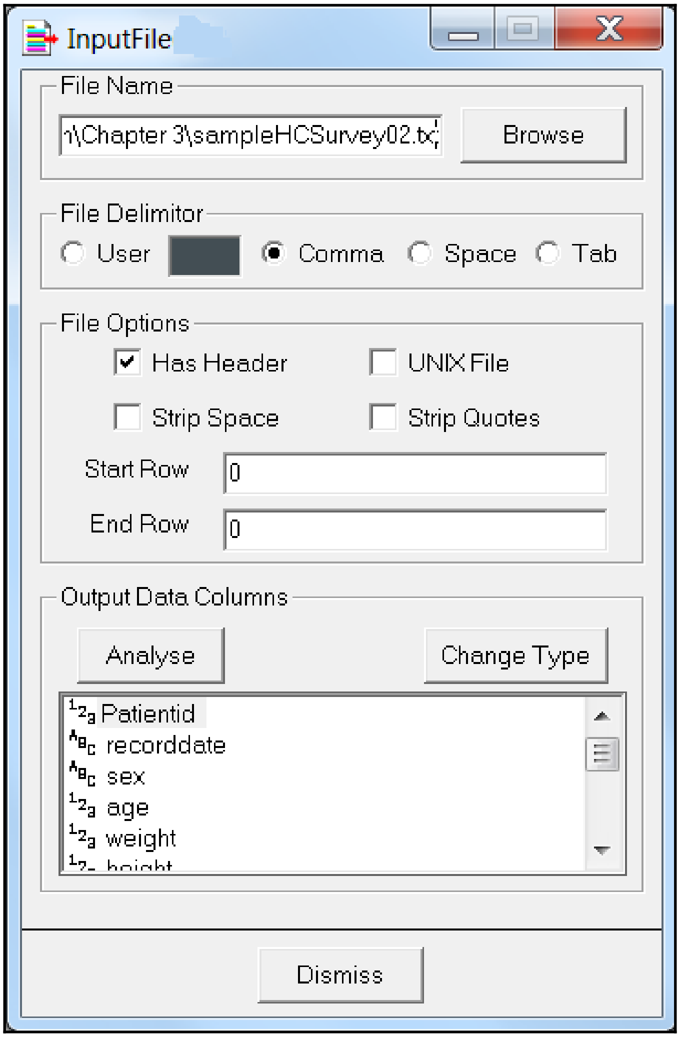

RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help

```
 recorddate
01
03
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
> 
```


DataManager
File
Input Nodes
Work Nodes
Output Nodes
OutputFile
OutputODBC
Quality
Statistics
Table
XYPlot
Execute
Help
1
2
3
Scene1
Scene2
Scene3
Scene4
Scene5
Scene6
0 %
11:00 AM
10/17/2016


1
Name Node
Connect Node
Edit Node
Delete Node
Validate Node
Execute Node
Cancel





InputFile

File Name

\Chapter 3\sampleHCSurvey02.tx

Browse

File Delimitor

○ User ● Comma ○ Space ○ Tab

File Options

☑ Has Header □ UNIX File

□ Strip Space □ Strip Quotes

Start Row 0

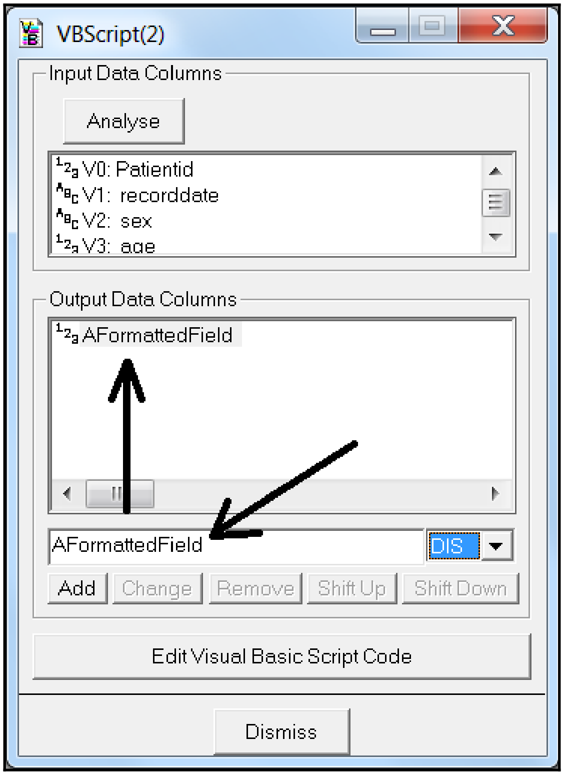
End Row 0

Output Data Columns

Analyse Change Type

Patientid
recorddate
sex
age
weight
height

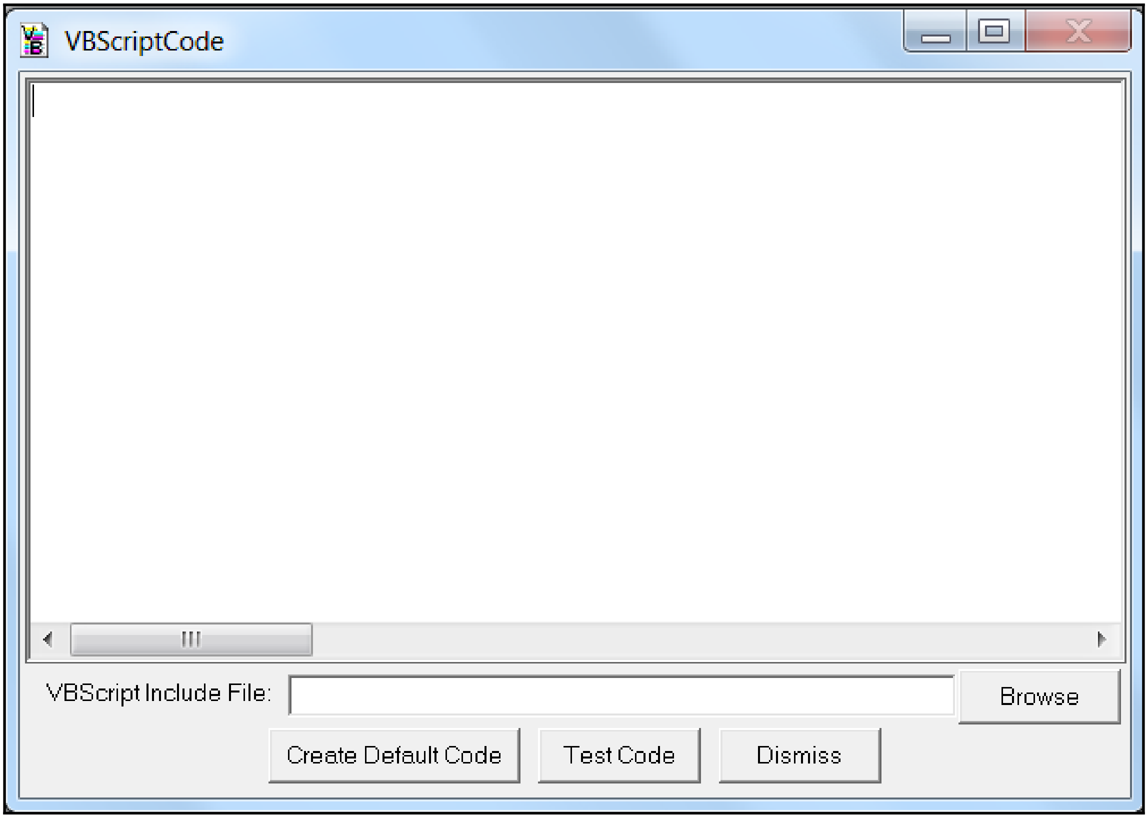
Dismiss

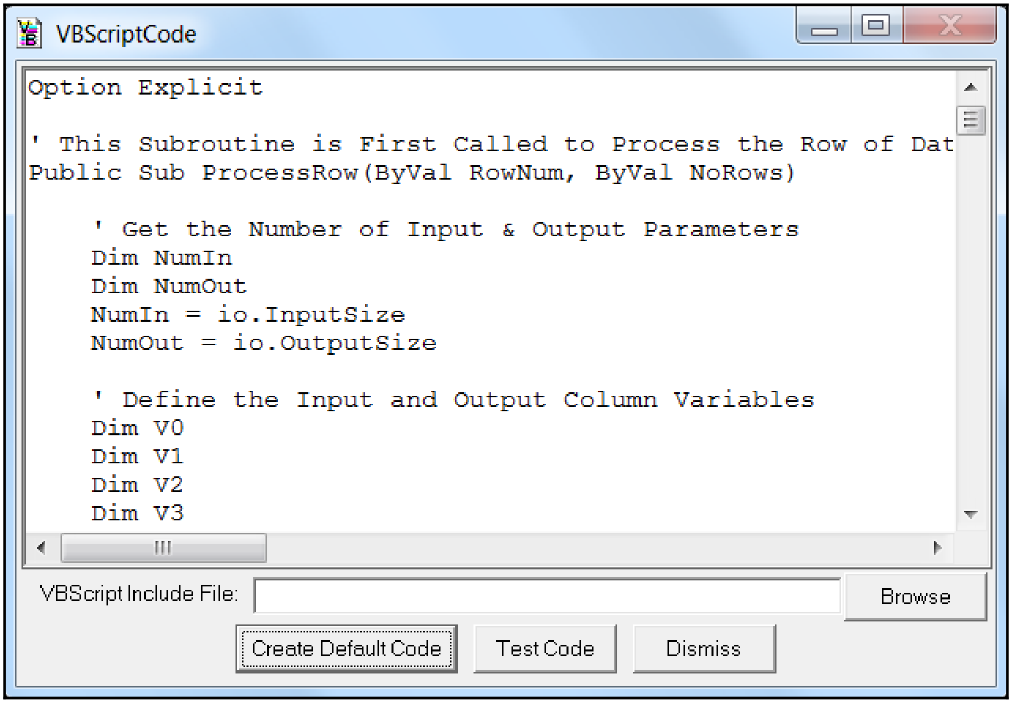


V0: Patientid
V1: recorddate
V2: sex
V3: age

AFormattedField

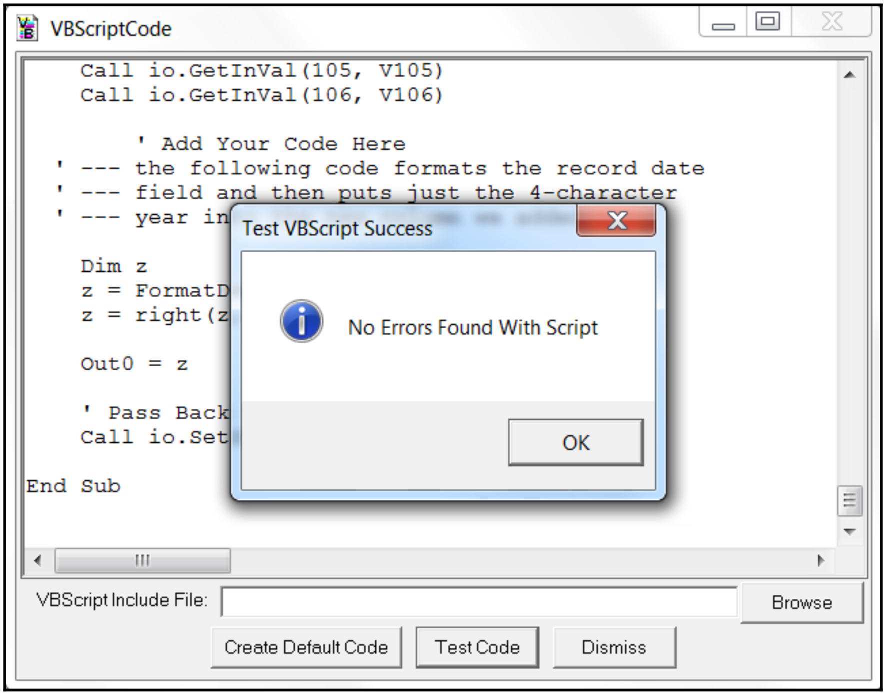
AFormattedField DIS


VBScriptCode
VBScript Include File:
Browse
Create Default Code
Test Code
Dismiss



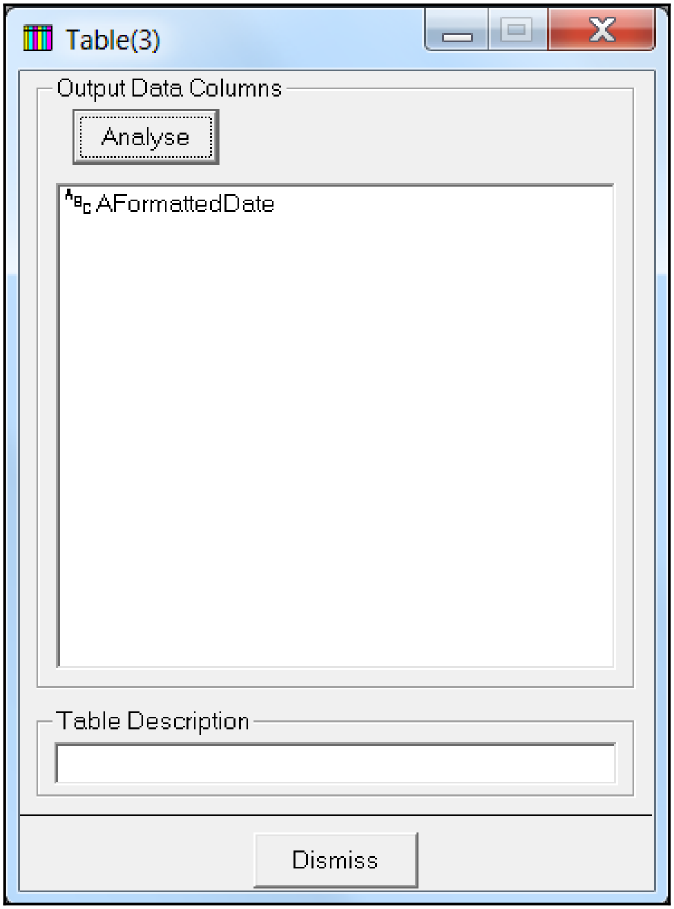
```
Option Explicit

' This Subroutine is First Called to Process the Row of Dat
Public Sub ProcessRow(ByVal RowNum, ByVal NoRows)

    ' Get the Number of Input & Output Parameters
    Dim NumIn
    Dim NumOut
    NumIn = io.InputSize
    NumOut = io.OutputSize

    ' Define the Input and Output Column Variables
    Dim V0
    Dim V1
    Dim V2
    Dim V3
```

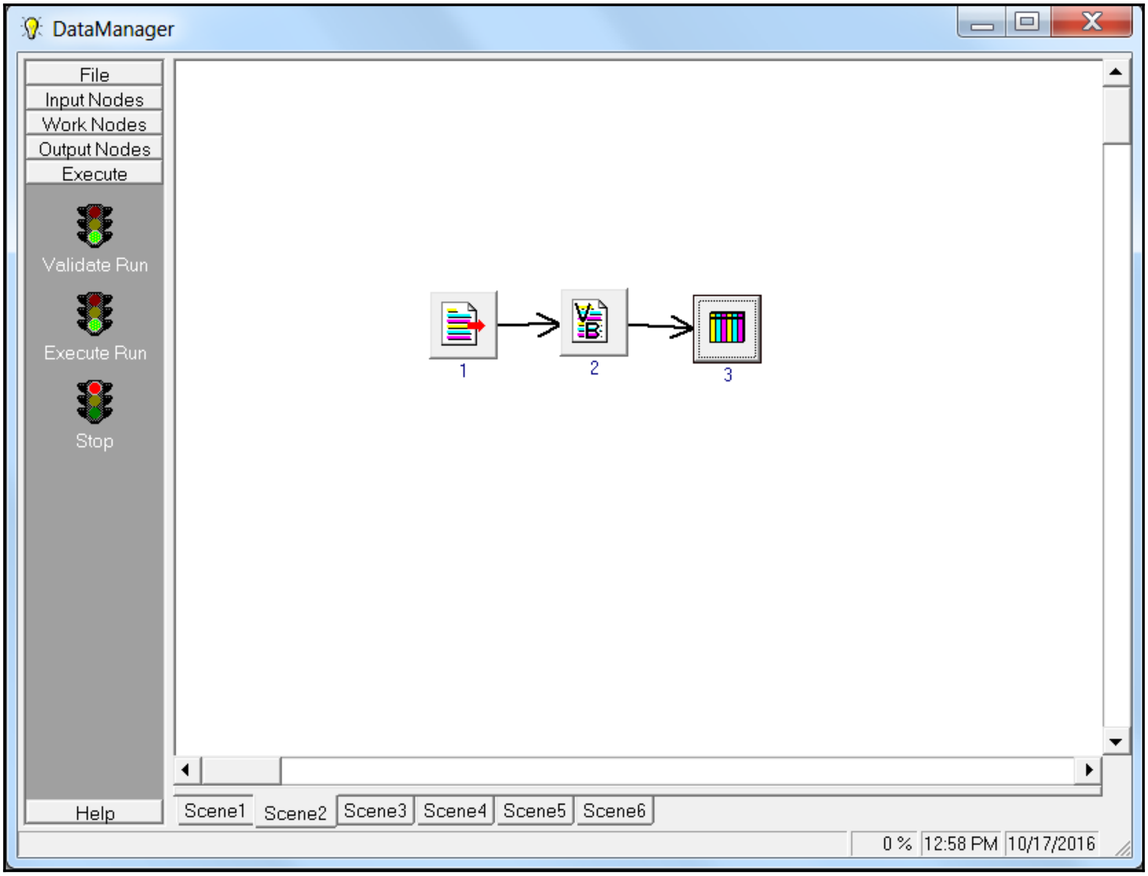
```
      ' Add Your Code Here
' --- the following code formats the record date
' --- field and then puts just the 4-character
' --- year into the new column we added

  Dim z
  z = FormatDateTime(V1,2)
  z = right(z,4)

  Out0 = z
```

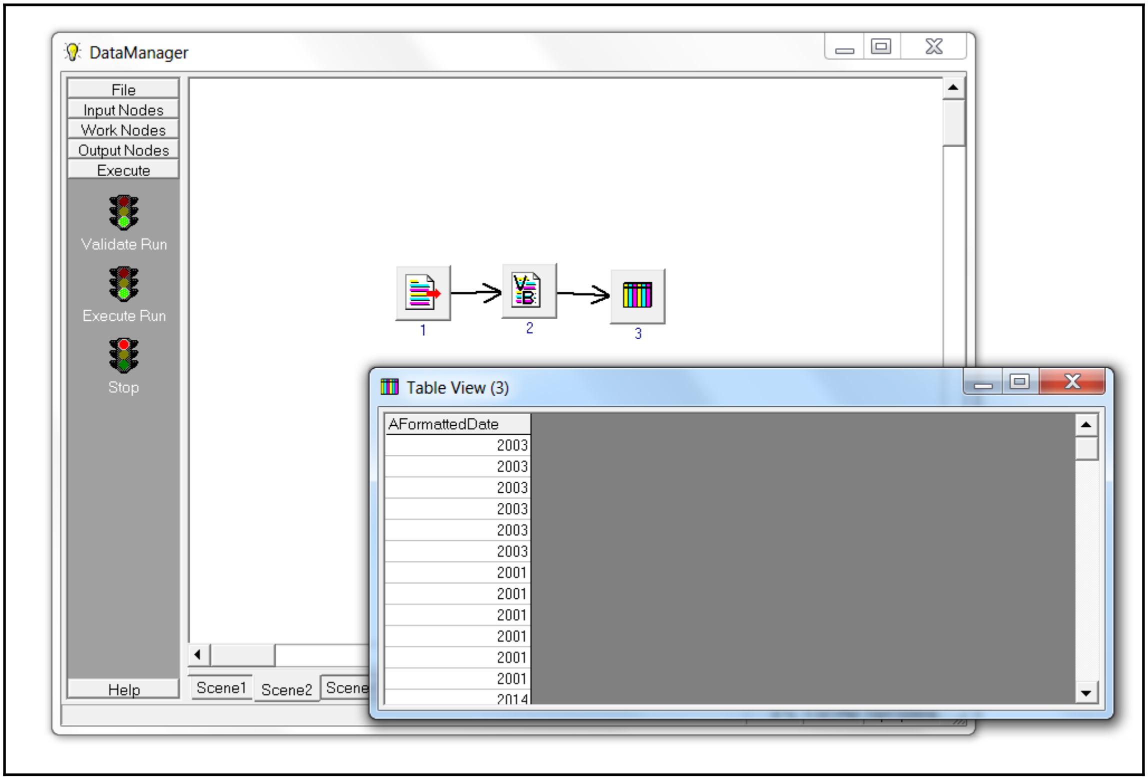


```
    Call io.GetInVal(105, V105)
    Call io.GetInVal(106, V106)

        ' Add Your Code Here
  ' --- the following code formats the record date
  ' --- field and then puts just the 4-character
  ' --- year in

    Dim z
    z = FormatD
    z = right(z

    Out0 = z

    ' Pass Back
    Call io.Set

End Sub
```

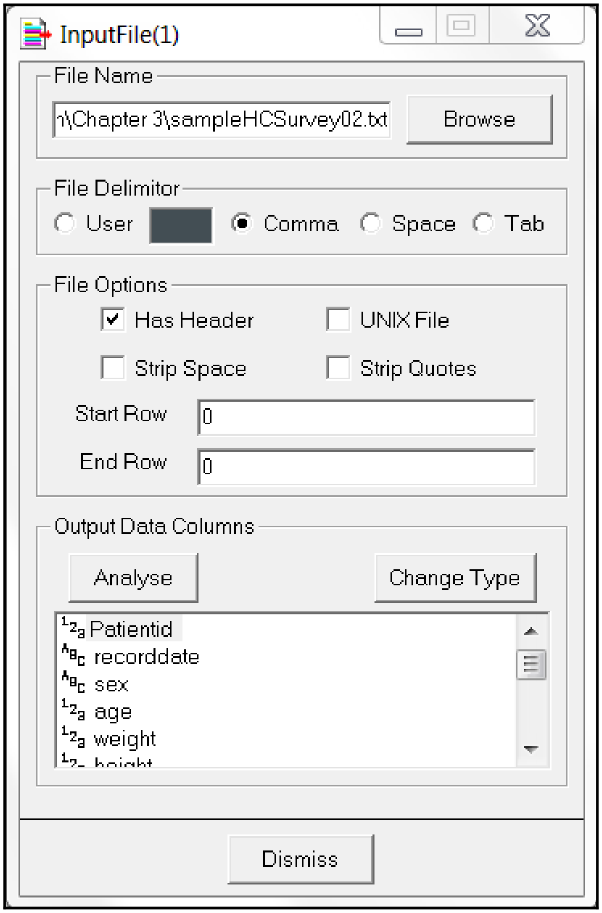
No Errors Found With Script


Table(3)
Output Data Columns
Analyse
AFormattedDate
Table Description
Dismiss


DataManager
File
Input Nodes
Work Nodes
Output Nodes
Execute
Validate Run
Execute Run
Stop
Help
1
2
3
Scene1
Scene2
Scene3
Scene4
Scene5
Scene6
0 %
12:58 PM
10/17/2016



AFormattedDate
2003
2003
2003
2003
2003
2003
2001
2001
2001
2001
2001
2001
2014



File Name

\Chapter 3\sampleHCSurvey02.txt

File Delimitor

○ User ● Comma ○ Space ○ Tab

File Options

☑ Has Header □ UNIX File

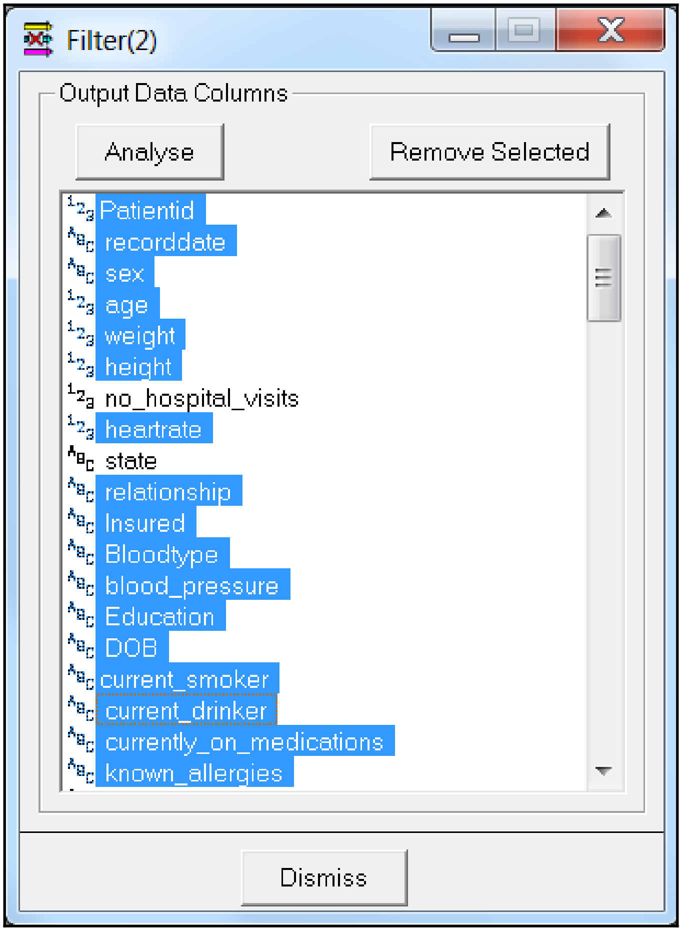
□ Strip Space □ Strip Quotes

Start Row 0

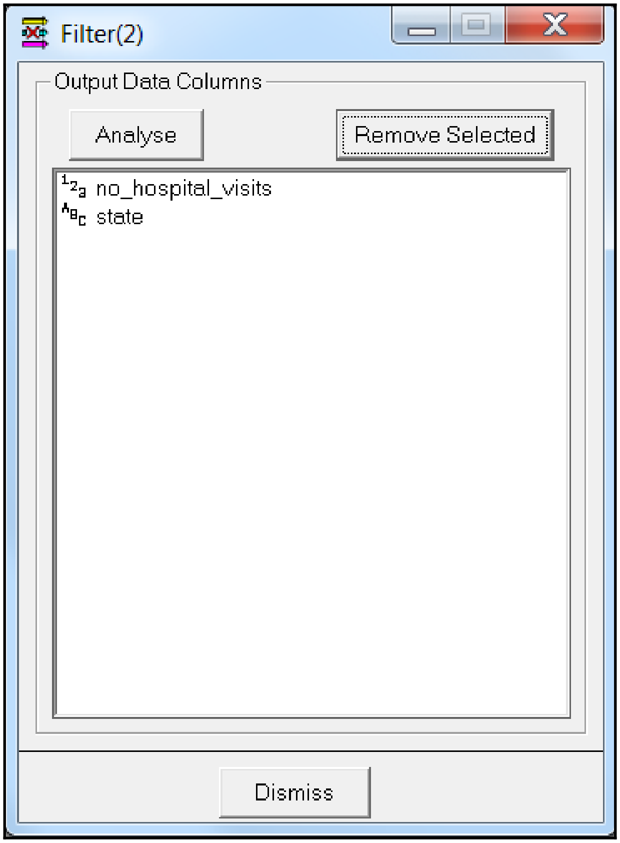
End Row 0

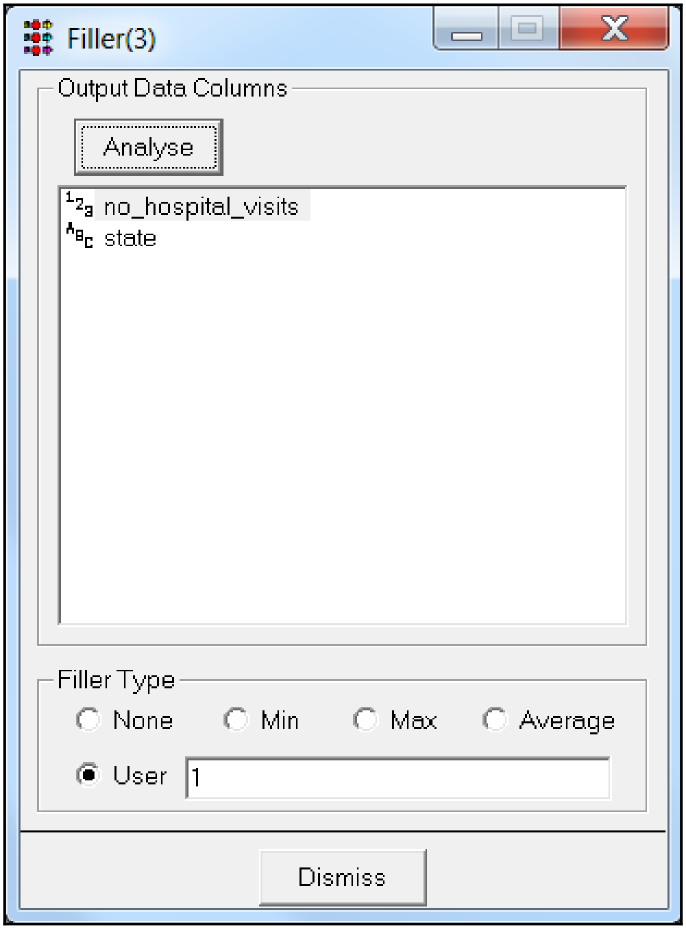
Output Data Columns

- 123 Patientid
- ABC recorddate
- ABC sex
- 123 age
- 123 weight
- 123 height



Patientid
recorddate
sex
age
weight
height
no_hospital_visits
heartrate
state
relationship
Insured
Bloodtype
blood_pressure
Education
DOB
current_smoker
current_drinker
currently_on_medications
known_allergies


Filter(2)
Output Data Columns
Analyse
Remove Selected
no_hospital_visits
state
Dismiss



Filler(3)

Output Data Columns

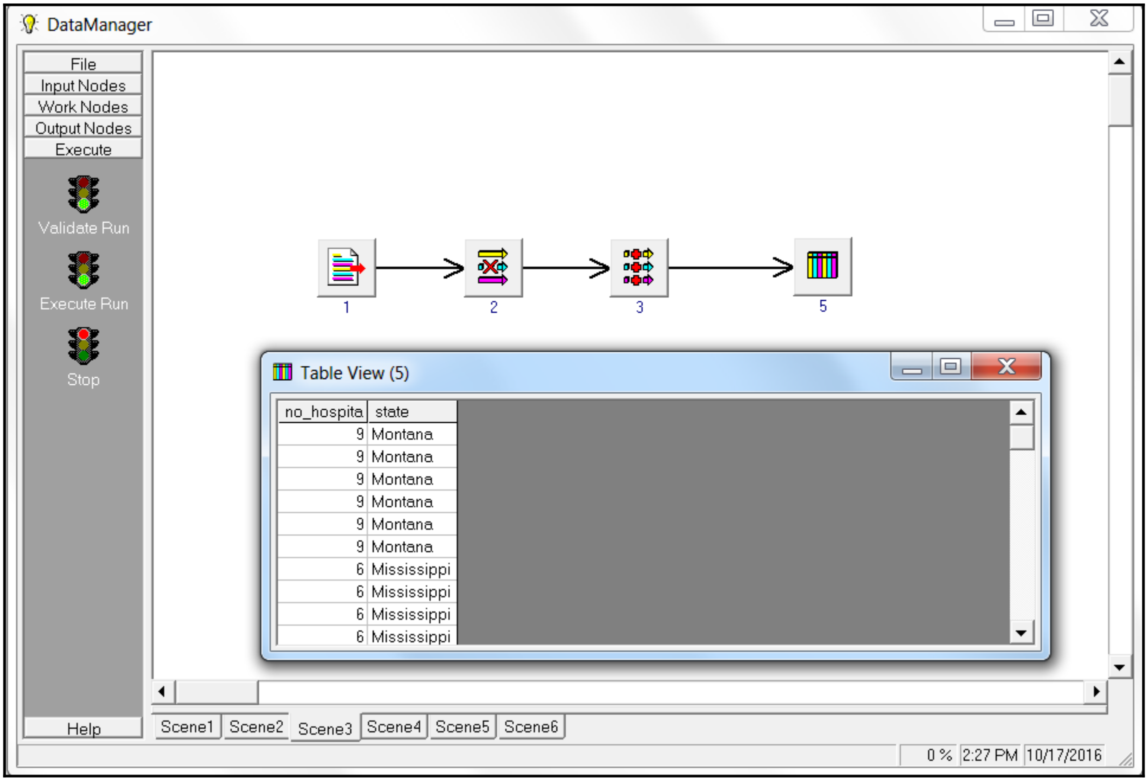
no_hospital_visits

state

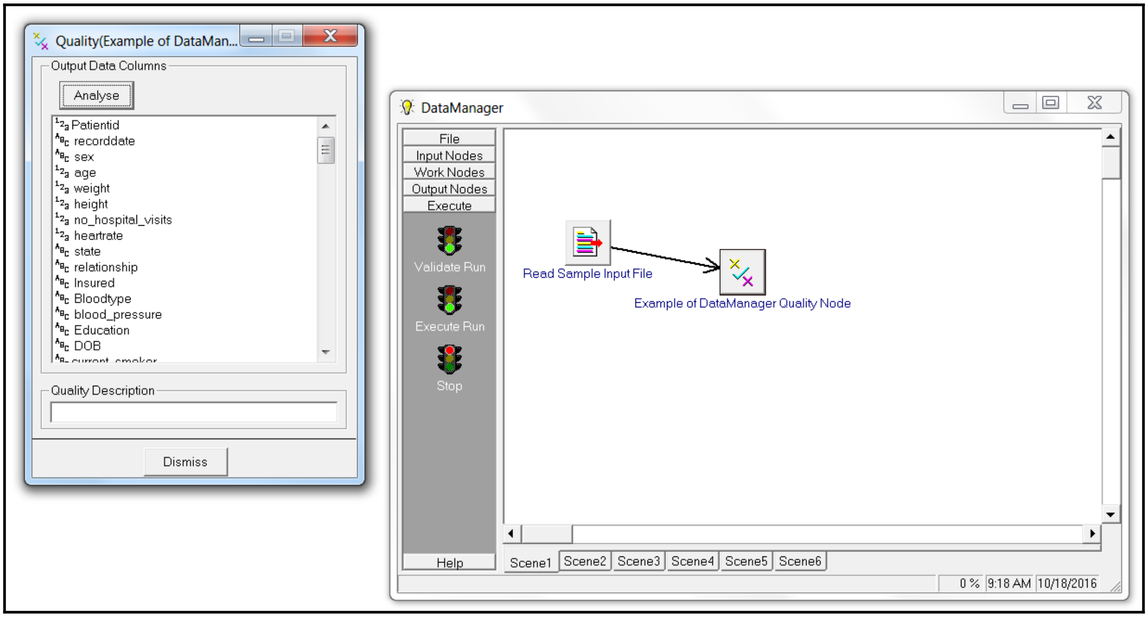
Filler Type

○ None ○ Min ○ Max ○ Average

● User 1



no_hospita	state
9	Montana
9	Montana
9	Montana
9	Montana
9	Montana
9	Montana
6	Mississippi
6	Mississippi
6	Mississippi
6	Mississippi



Quality(Example of DataMan...

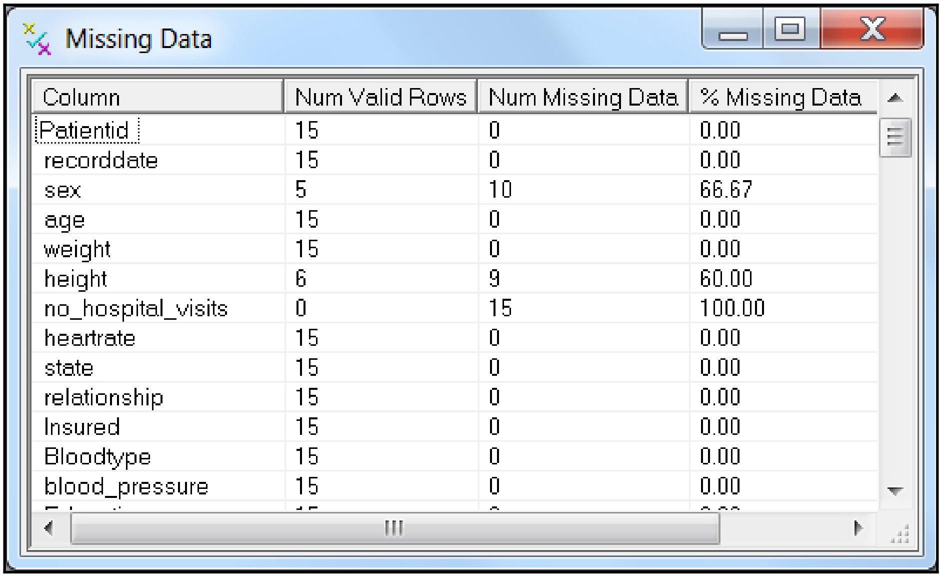
Output Data Columns

Analyse

- Patientid
- recorddate
- sex
- age
- weight
- height
- no_hospital_visits
- heartrate
- state
- relationship
- Insured
- Bloodtype
- blood_pressure
- Education
- DOB
- current_smoker

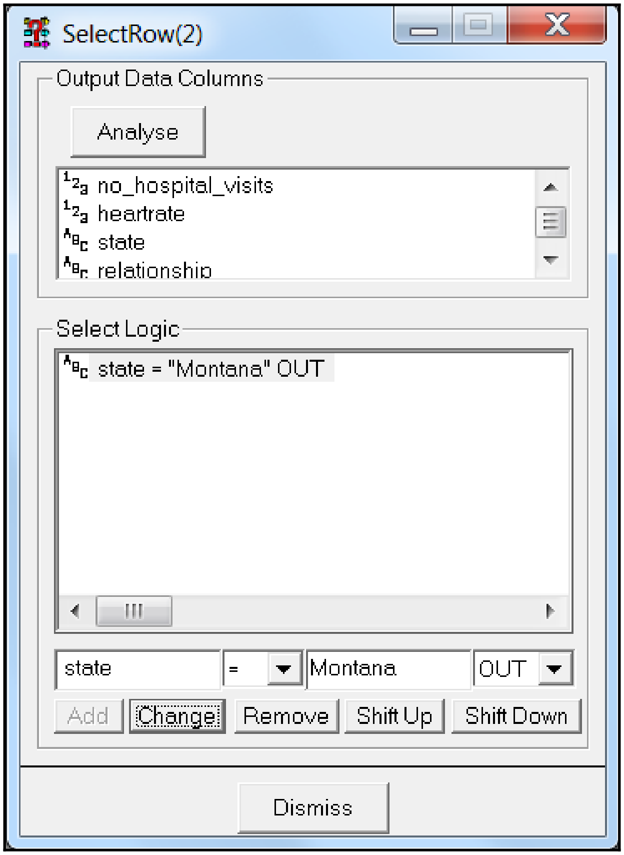
Quality Description

Dismiss



Missing Data

Column	Num Valid Rows	Num Missing Data	% Missing Data
Patientid	15	0	0.00
recorddate	15	0	0.00
sex	5	10	66.67
age	15	0	0.00
weight	15	0	0.00
height	6	9	60.00
no_hospital_visits	0	15	100.00
heartrate	15	0	0.00
state	15	0	0.00
relationship	15	0	0.00
Insured	15	0	0.00
Bloodtype	15	0	0.00
blood_pressure	15	0	0.00



SelectRow(2)

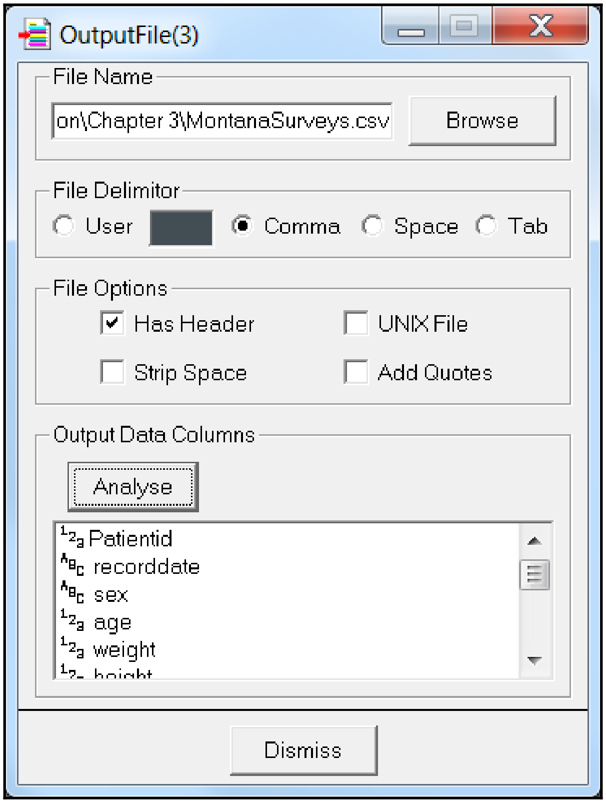
Output Data Columns

no_hospital_visits
heartrate
state
relationship

Select Logic

state = "Montana" OUT

state = Montana OUT



File Name: on\Chapter 3\MontanaSurveys.csv

File Delimitor: ○ User ● Comma ○ Space ○ Tab

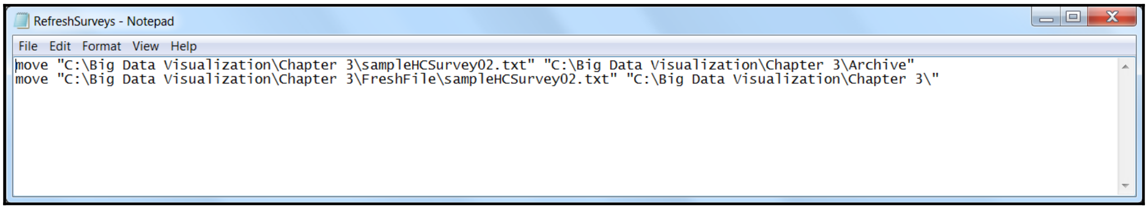
File Options:

☑ Has Header ☐ UNIX File

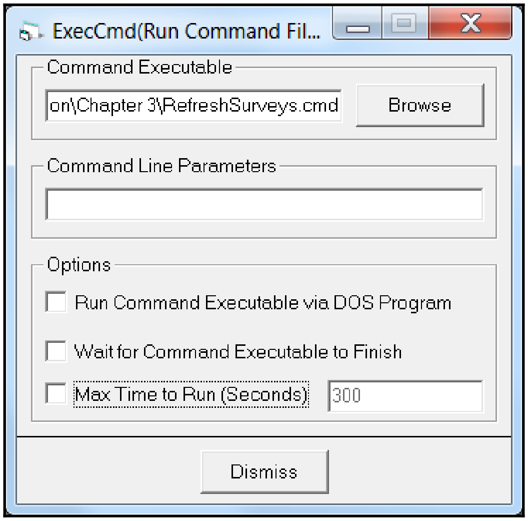
☐ Strip Space ☐ Add Quotes

Output Data Columns

Patientid
recorddate
sex
age
weight
height



```
move "C:\Big Data Visualization\Chapter 3\sampleHCSurvey02.txt" "C:\Big Data Visualization\Chapter 3\Archive"
move "C:\Big Data Visualization\Chapter 3\FreshFile\sampleHCSurvey02.txt" "C:\Big Data Visualization\Chapter 3\"
```



ExecCmd(Run Command Fil...

Command Executable

on\Chapter 3\RefreshSurveys.cmd

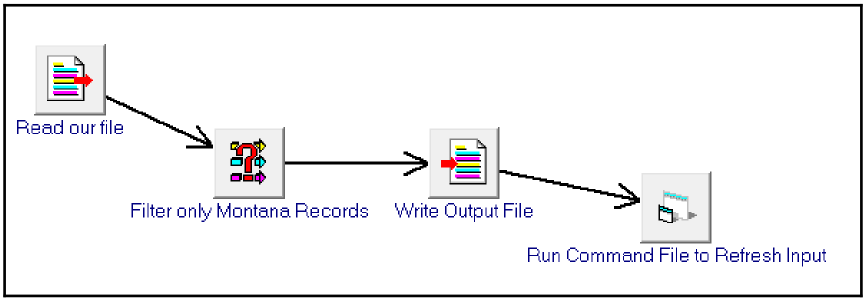
Command Line Parameters

Options

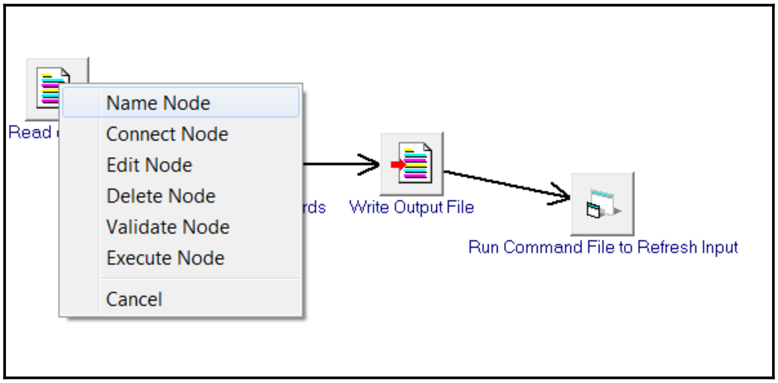
□ Run Command Executable via DOS Program

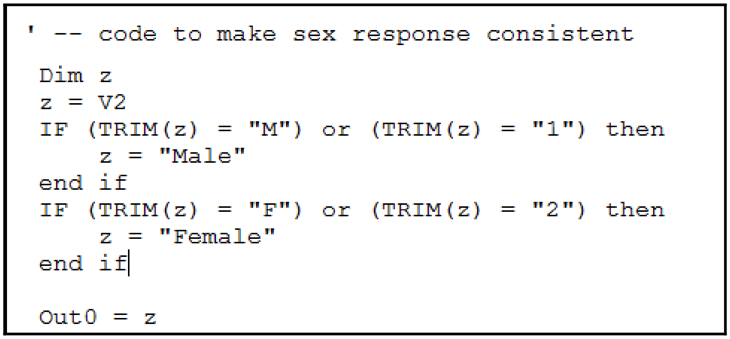
□ Wait for Command Executable to Finish

□ Max Time to Run (Seconds) 300

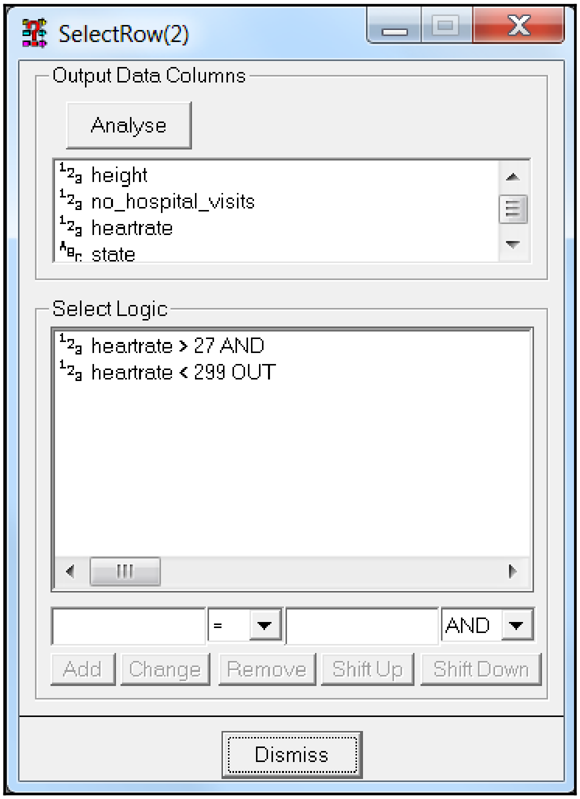


Read our file
Filter only Montana Records
Write Output File
Run Command File to Refresh Input

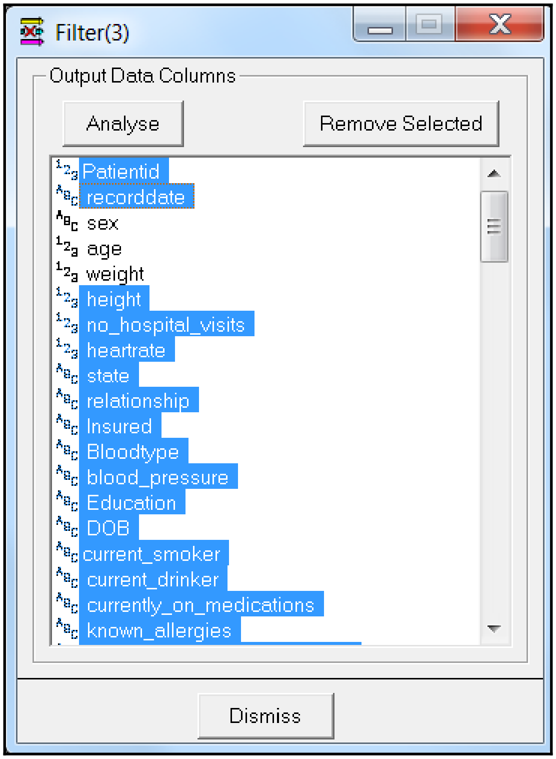




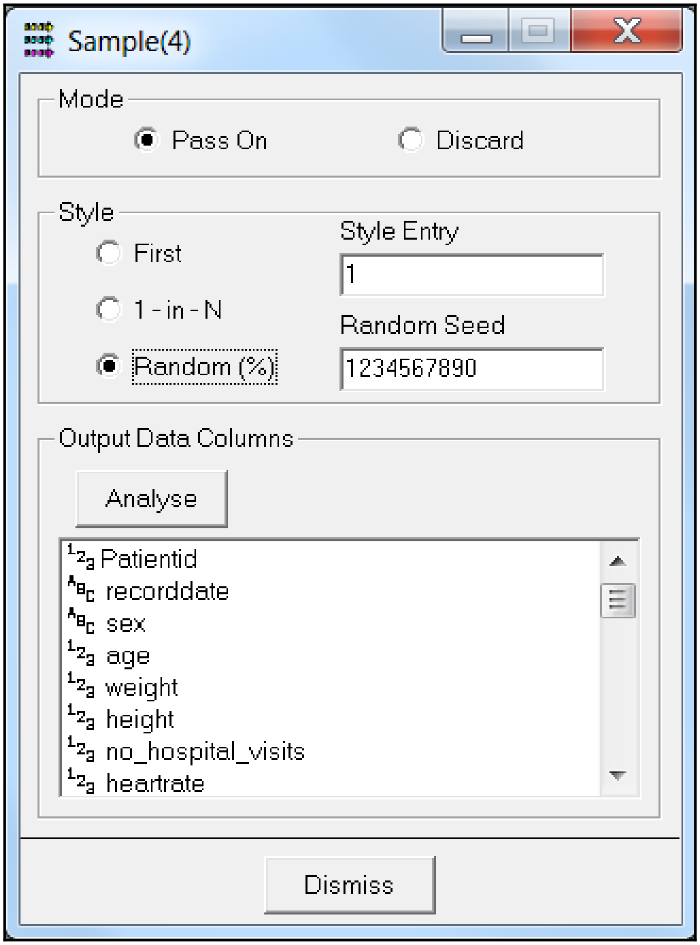
```
' -- code to make sex response consistent

 Dim z
 z = V2
 IF (TRIM(z) = "M") or (TRIM(z) = "1") then
     z = "Male"
 end if
 IF (TRIM(z) = "F") or (TRIM(z) = "2") then
     z = "Female"
 end if

 Out0 = z
```



heartrate > 27 AND
heartrate < 299 OUT



Patientid
recorddate
sex
age
weight
height
no_hospital_visits
heartrate
state
relationship
Insured
Bloodtype
blood_pressure
Education
DOB
current_smoker
current_drinker
currently_on_medications
known_allergies



Mode

● Pass On ○ Discard

Style

○ First

○ 1 - in - N

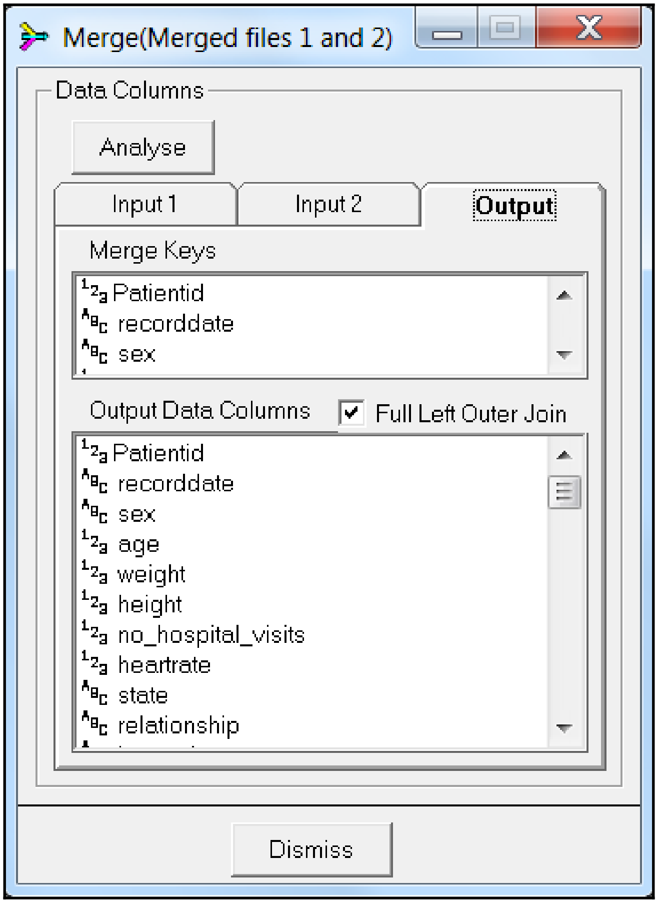
● Random (%)

Style Entry: 1

Random Seed: 1234567890

Output Data Columns

Patientid
recorddate
sex
age
weight
height
no_hospital_visits
heartrate

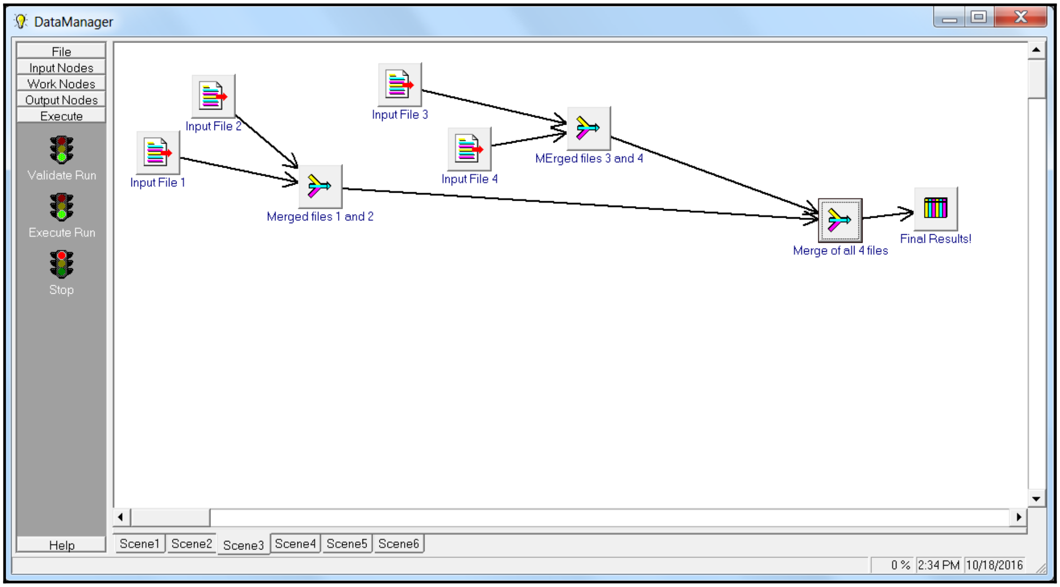


Merge Keys

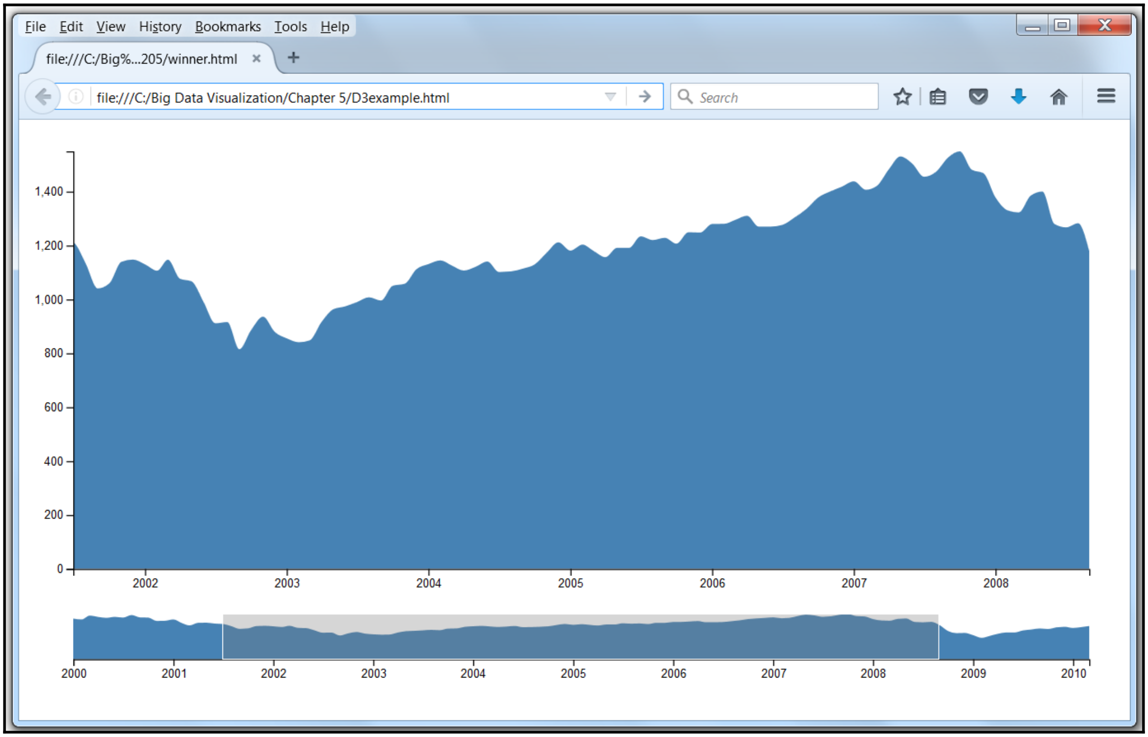
- Patientid
- recorddate
- sex

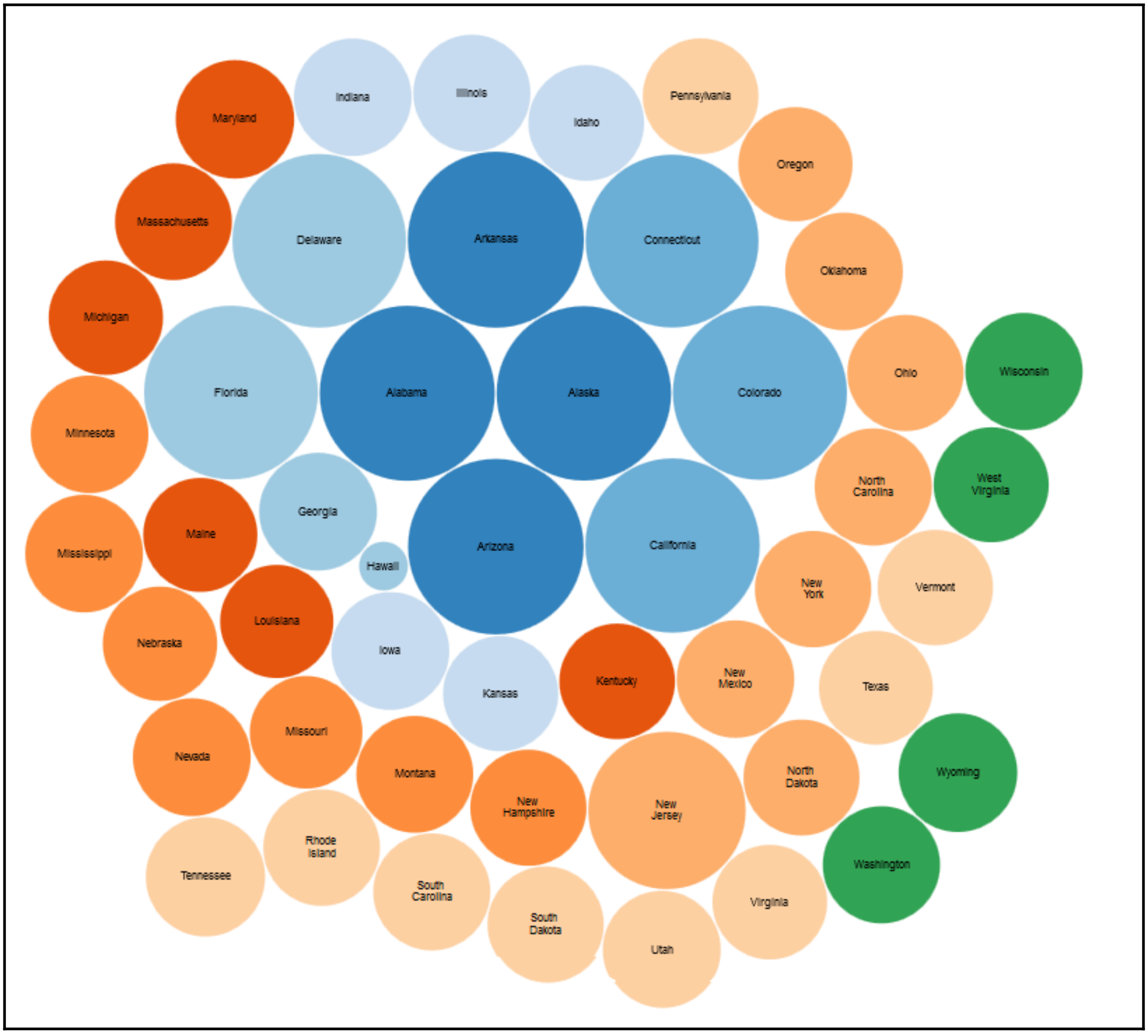
Output Data Columns ☑ Full Left Outer Join

- Patientid
- recorddate
- sex
- age
- weight
- height
- no_hospital_visits
- heartrate
- state
- relationship

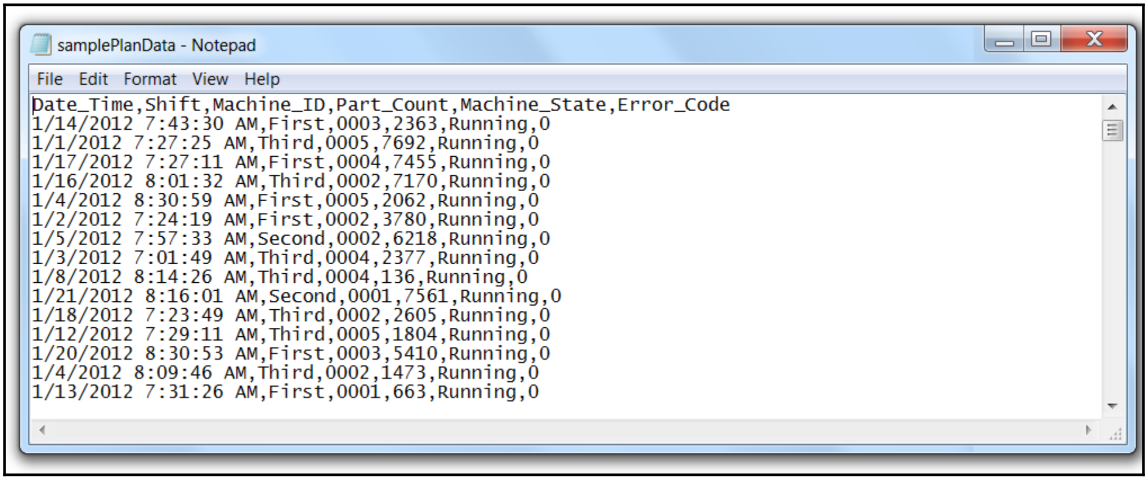

DataManager
File
Input Nodes
Work Nodes
Output Nodes
Execute
Validate Run
Execute Run
Stop
Help
Input File 2
Input File 1
Merged files 1 and 2
Input File 3
Input File 4
MErged files 3 and 4
Merge of all 4 files
Final Results!
Scene1
Scene2
Scene3
Scene4
Scene5
Scene6
0 %
2:34 PM
10/18/2016

Chapter 5: Displaying Results Using D3

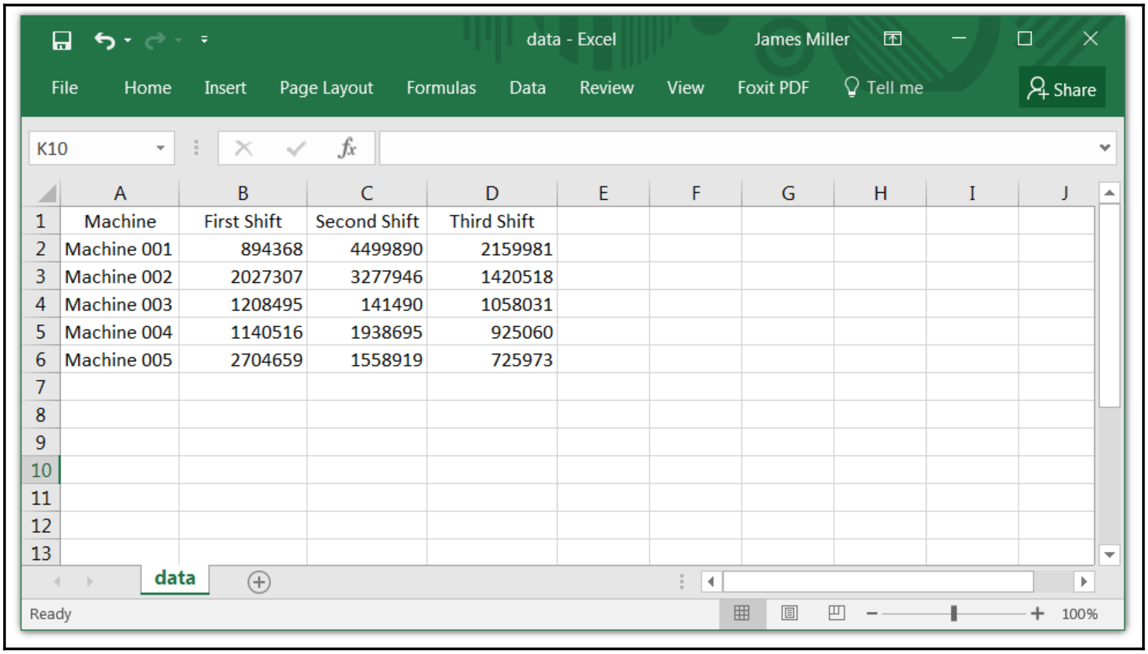

File Edit View History Bookmarks Tools Help
file:///C:/Big%...205/winner.html
file:///C:/Big Data Visualization/Chapter 5/D3example.html
Search
1,400
1,200
1,000
800
600
400
200
0
2002
2003
2004
2005
2006
2007
2008
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010



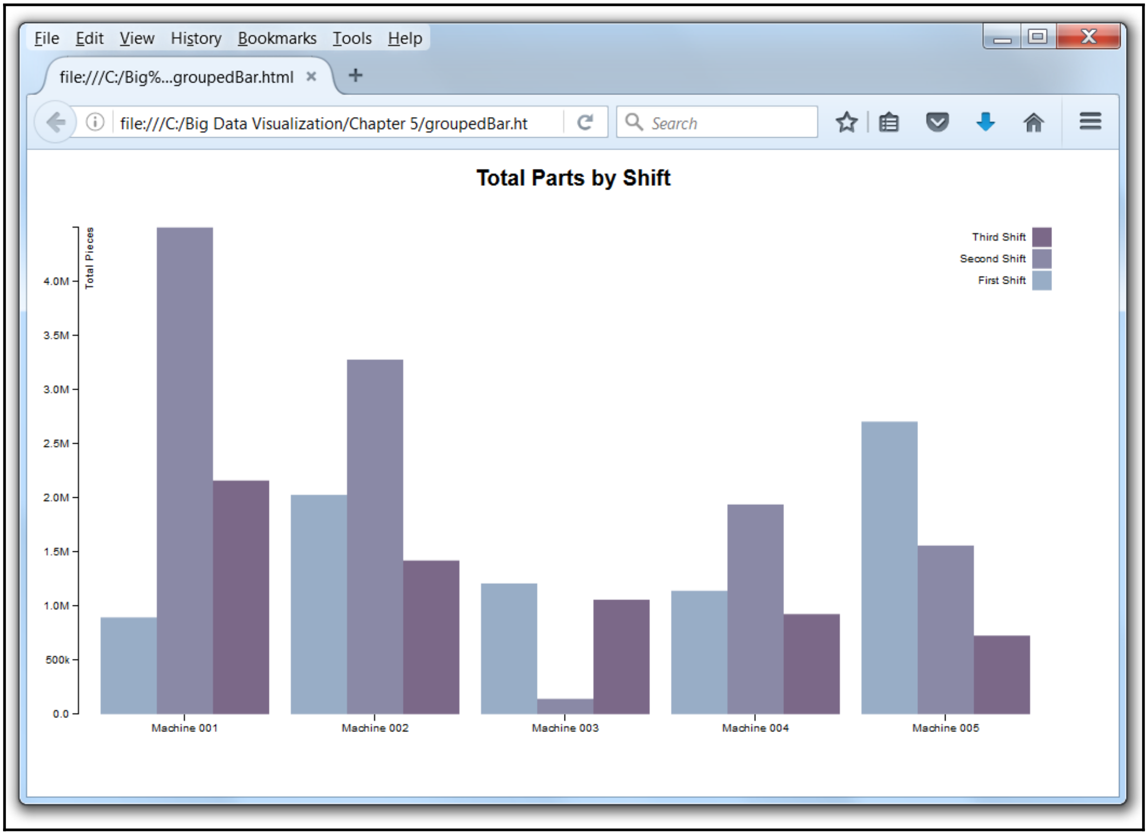
Maryland
Indiana
Illinois
Idaho
Pennsylvania
Oregon
Massachusetts
Delaware
Arkansas
Connecticut
Oklahoma
Michigan
Florida
Alabama
Alaska
Colorado
Ohio
Wisconsin
Minnesota
North Carolina
West Virginia
Georgia
Maine
Mississippi
Hawaii
Arizona
California
New York
Vermont
Louisiana
Nebraska
Iowa
Kansas
Kentucky
New Mexico
Texas
Missouri
Nevada
Montana
New Hampshire
New Jersey
North Dakota
Wyoming
Rhode Island
Tennessee
South Carolina
South Dakota
Utah
Virginia
Washington

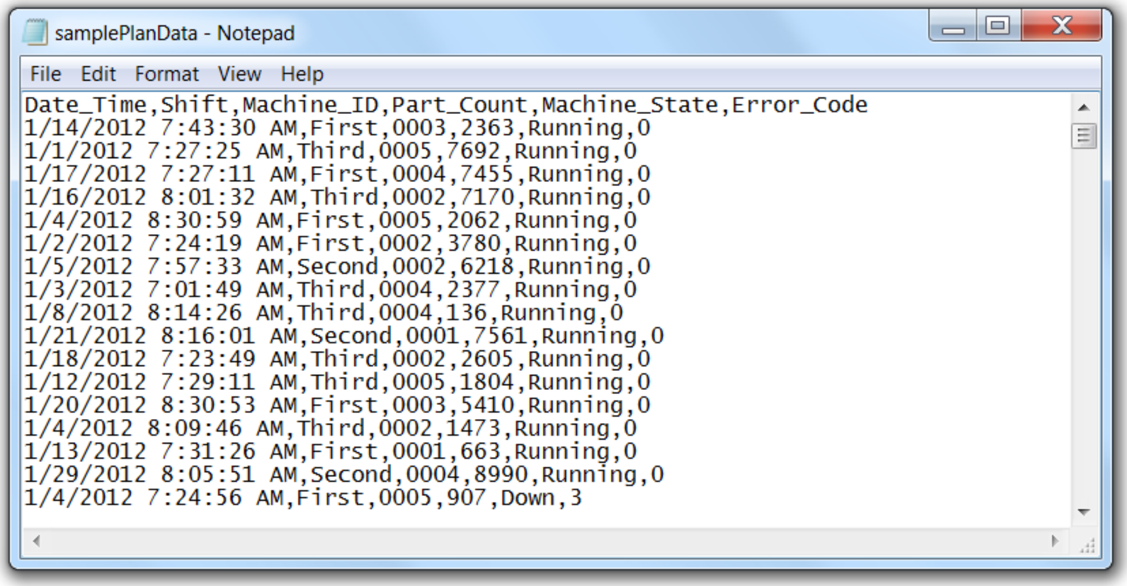


```
Date_Time,Shift,Machine_ID,Part_Count,Machine_State,Error_Code
1/14/2012 7:43:30 AM,First,0003,2363,Running,0
1/1/2012 7:27:25 AM,Third,0005,7692,Running,0
1/17/2012 7:27:11 AM,First,0004,7455,Running,0
1/16/2012 8:01:32 AM,Third,0002,7170,Running,0
1/4/2012 8:30:59 AM,First,0005,2062,Running,0
1/2/2012 7:24:19 AM,First,0002,3780,Running,0
1/5/2012 7:57:33 AM,Second,0002,6218,Running,0
1/3/2012 7:01:49 AM,Third,0004,2377,Running,0
1/8/2012 8:14:26 AM,Third,0004,136,Running,0
1/21/2012 8:16:01 AM,Second,0001,7561,Running,0
1/18/2012 7:23:49 AM,Third,0002,2605,Running,0
1/12/2012 7:29:11 AM,Third,0005,1804,Running,0
1/20/2012 8:30:53 AM,First,0003,5410,Running,0
1/4/2012 8:09:46 AM,Third,0002,1473,Running,0
1/13/2012 7:31:26 AM,First,0001,663,Running,0
```

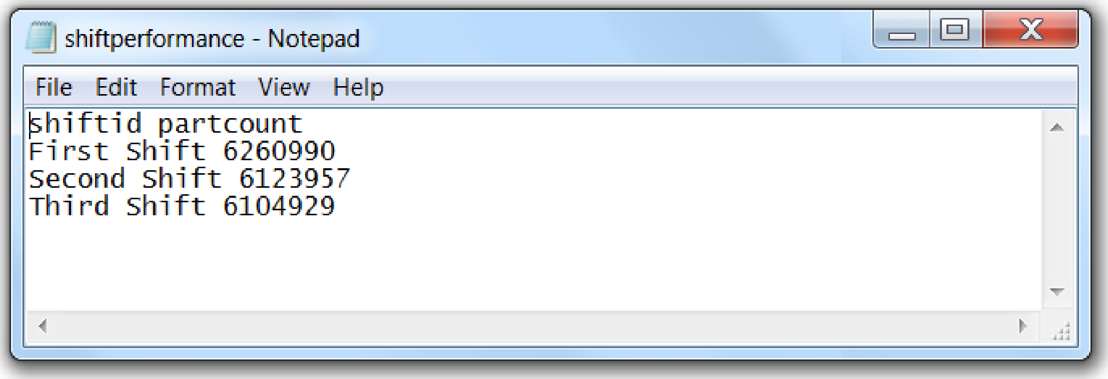


Machine	First Shift	Second Shift	Third Shift
Machine 001	894368	4499890	2159981
Machine 002	2027307	3277946	1420518
Machine 003	1208495	141490	1058031
Machine 004	1140516	1938695	925060
Machine 005	2704659	1558919	725973

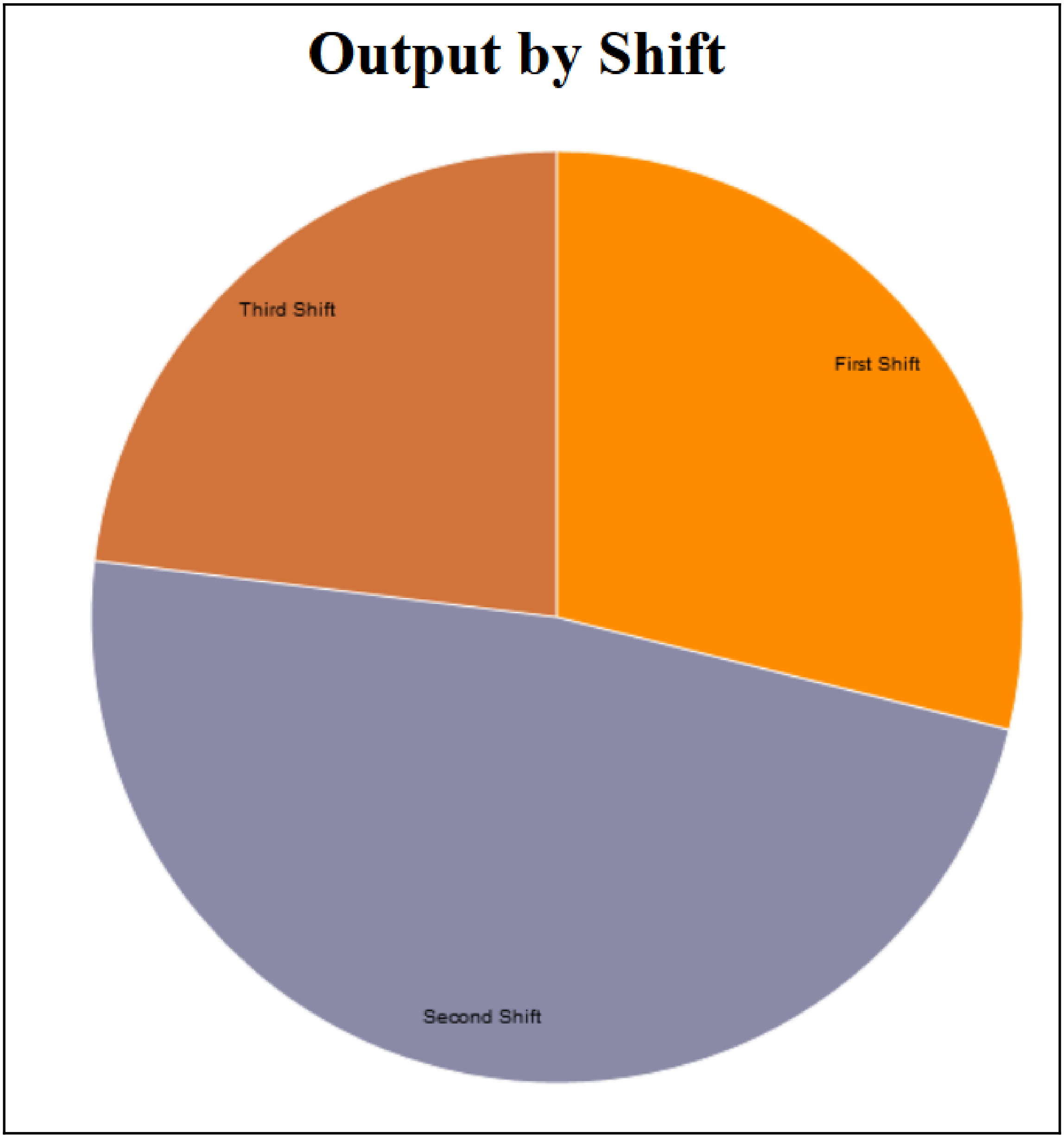

File
Edit
View
History
Bookmarks
Tools
Help
file:///C:/Big%...groupedBar.html
file:///C:/Big Data Visualization/Chapter 5/groupedBar.ht
Search
Total Parts by Shift
Total Pieces
4.0M
3.5M
3.0M
2.5M
2.0M
1.5M
1.0M
500k
0.0
Third Shift
Second Shift
First Shift
Machine 001
Machine 002
Machine 003
Machine 004
Machine 005



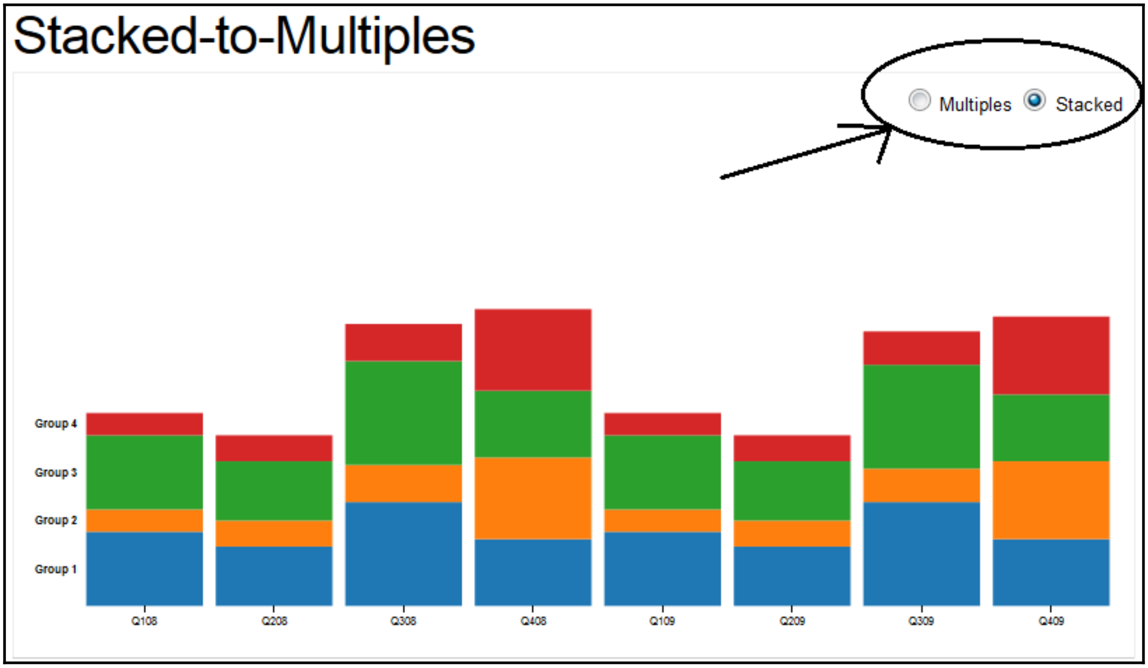
```
Date_Time,Shift,Machine_ID,Part_Count,Machine_State,Error_Code
1/14/2012 7:43:30 AM,First,0003,2363,Running,0
1/1/2012 7:27:25 AM,Third,0005,7692,Running,0
1/17/2012 7:27:11 AM,First,0004,7455,Running,0
1/16/2012 8:01:32 AM,Third,0002,7170,Running,0
1/4/2012 8:30:59 AM,First,0005,2062,Running,0
1/2/2012 7:24:19 AM,First,0002,3780,Running,0
1/5/2012 7:57:33 AM,Second,0002,6218,Running,0
1/3/2012 7:01:49 AM,Third,0004,2377,Running,0
1/8/2012 8:14:26 AM,Third,0004,136,Running,0
1/21/2012 8:16:01 AM,Second,0001,7561,Running,0
1/18/2012 7:23:49 AM,Third,0002,2605,Running,0
1/12/2012 7:29:11 AM,Third,0005,1804,Running,0
1/20/2012 8:30:53 AM,First,0003,5410,Running,0
1/4/2012 8:09:46 AM,Third,0002,1473,Running,0
1/13/2012 7:31:26 AM,First,0001,663,Running,0
1/29/2012 8:05:51 AM,Second,0004,8990,Running,0
1/4/2012 7:24:56 AM,First,0005,907,Down,3
```



```
shiftid partcount
First Shift 6260990
Second Shift 6123957
Third Shift 6104929
```

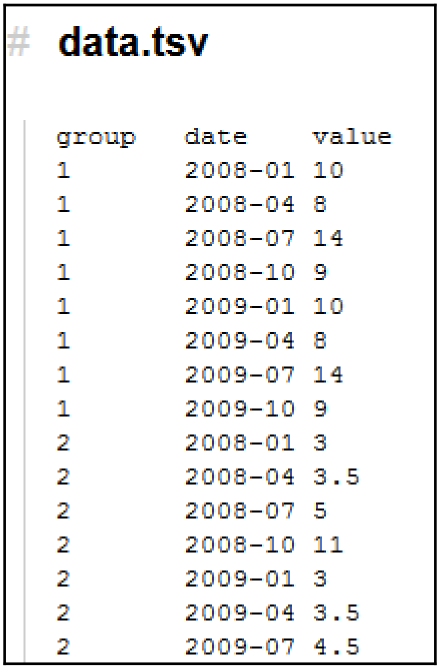


Output by Shift



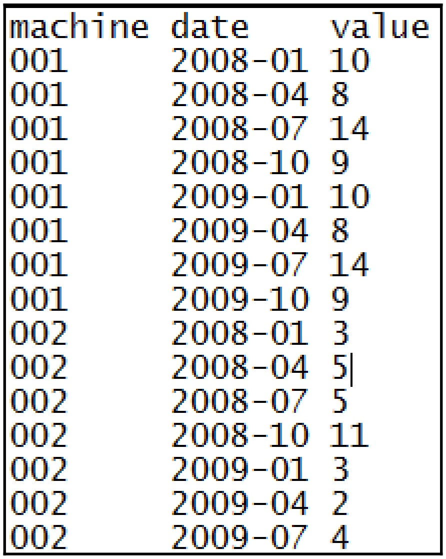
Stacked-to-Multiples

○ Multiples ● Stacked

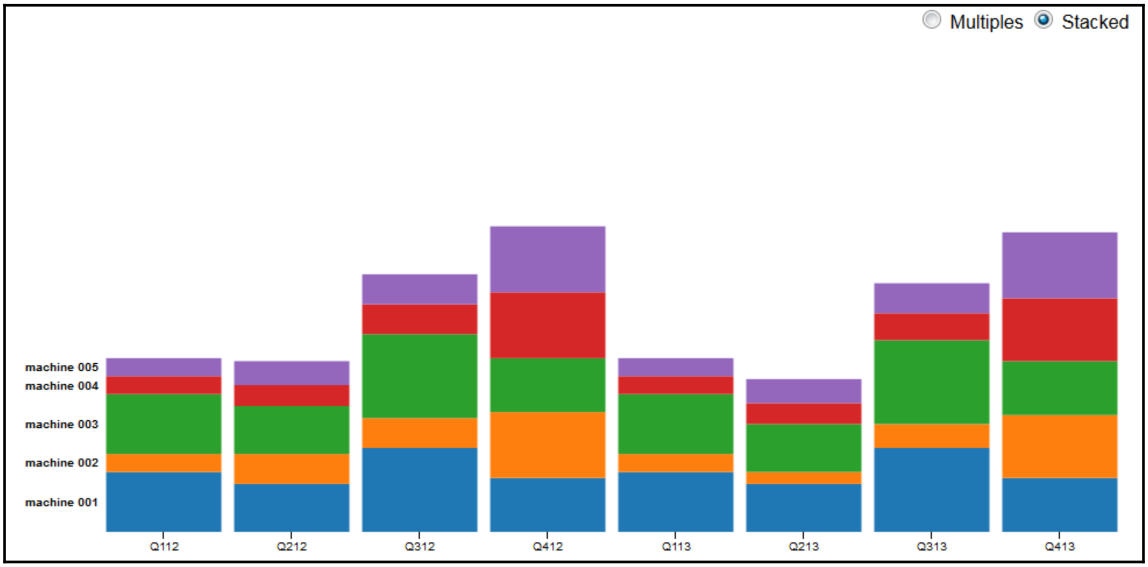


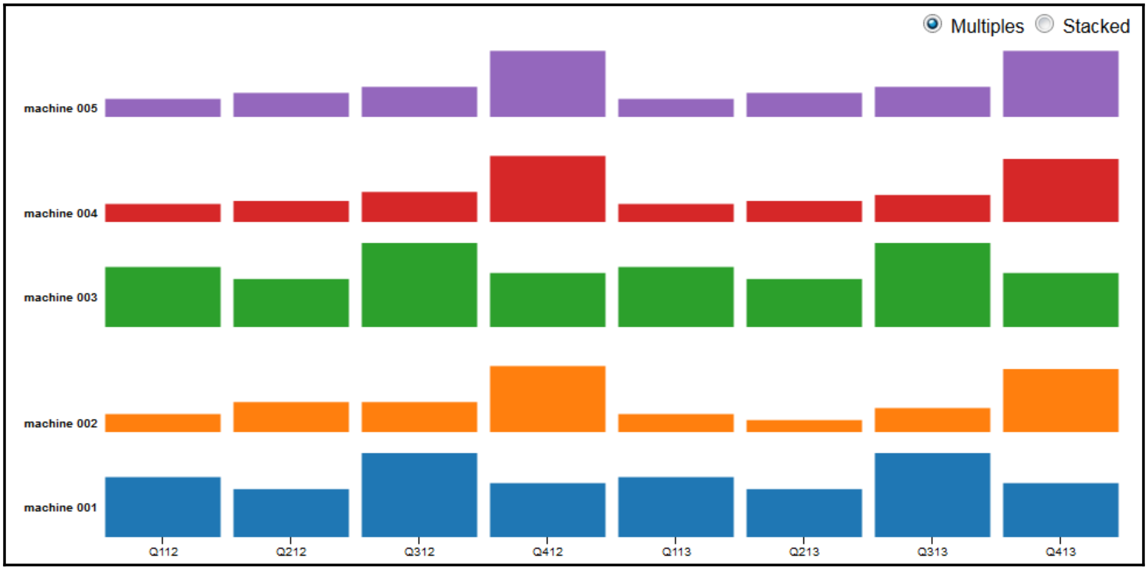
data.tsv

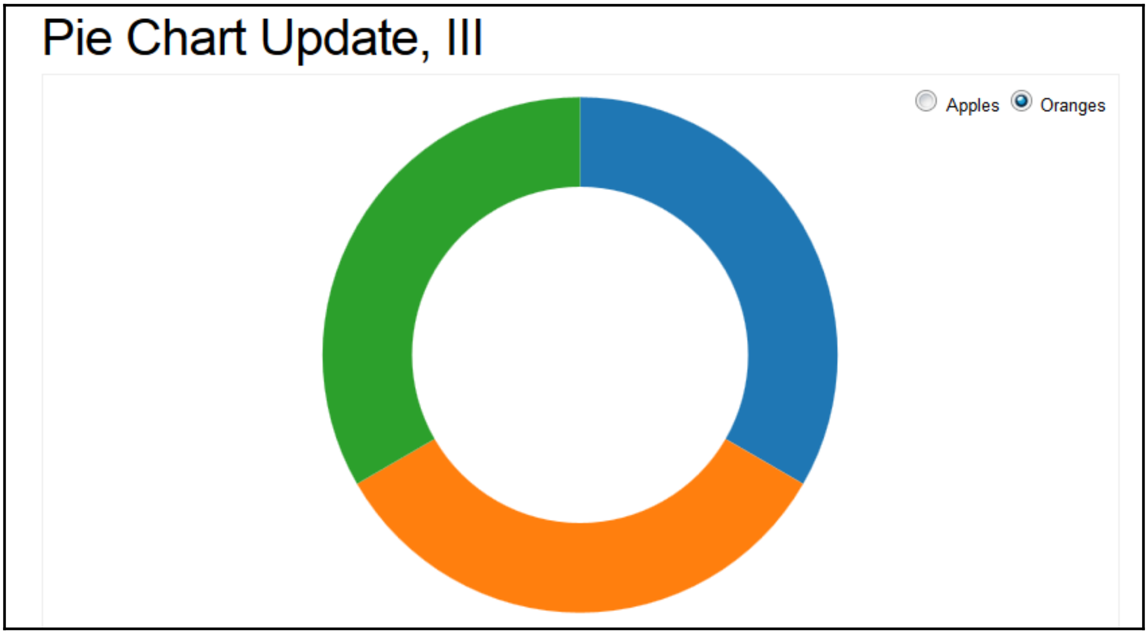
```
group    date    value
1        2008-01 10
1        2008-04 8
1        2008-07 14
1        2008-10 9
1        2009-01 10
1        2009-04 8
1        2009-07 14
1        2009-10 9
2        2008-01 3
2        2008-04 3.5
2        2008-07 5
2        2008-10 11
2        2009-01 3
2        2009-04 3.5
2        2009-07 4.5
```

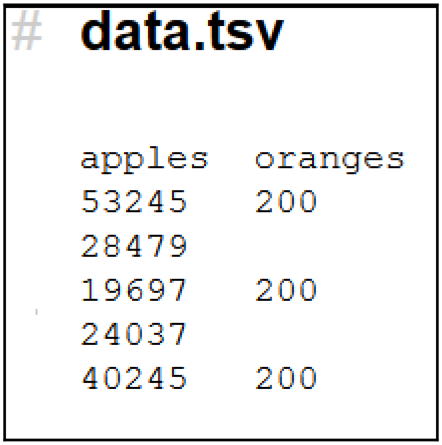


machine	date	value
001	2008-01	10
001	2008-04	8
001	2008-07	14
001	2008-10	9
001	2009-01	10
001	2009-04	8
001	2009-07	14
001	2009-10	9
002	2008-01	3
002	2008-04	5
002	2008-07	5
002	2008-10	11
002	2009-01	3
002	2009-04	2
002	2009-07	4



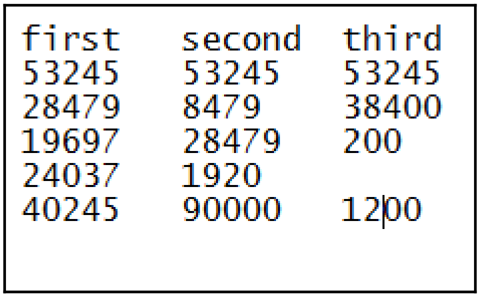



Pie Chart Update, III

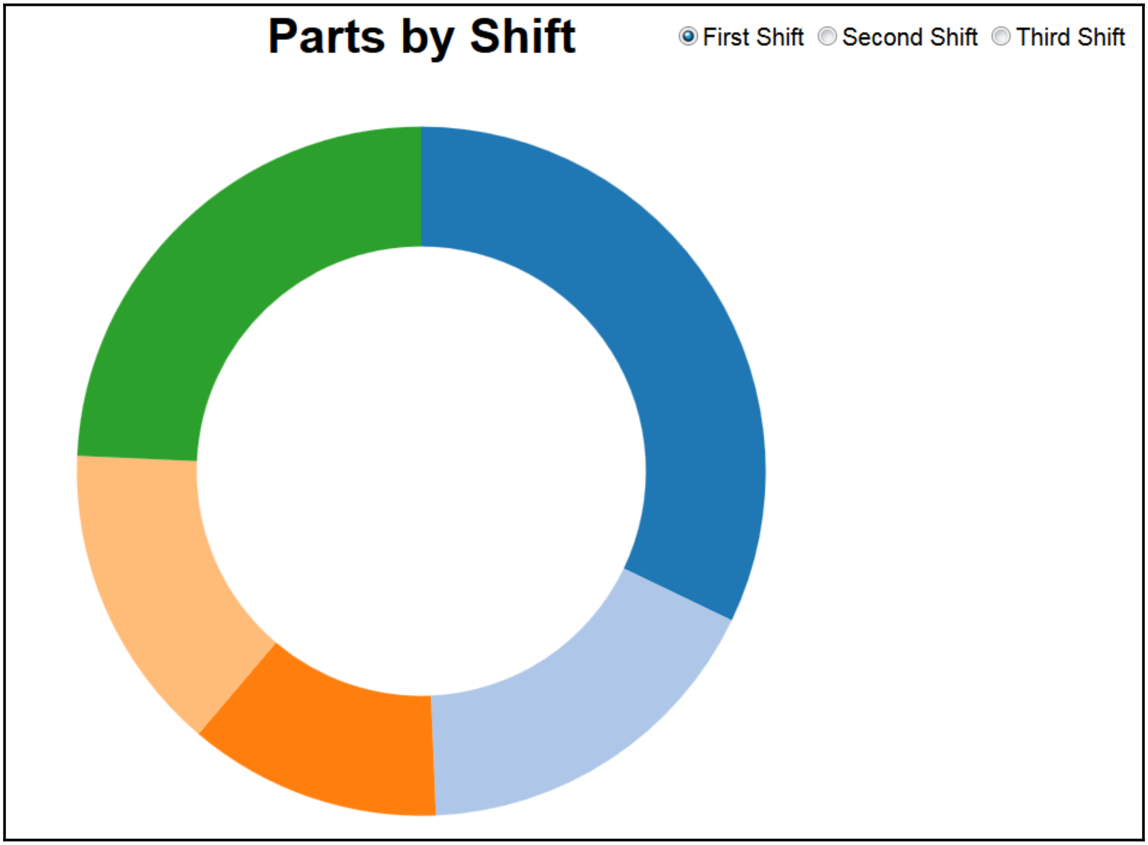


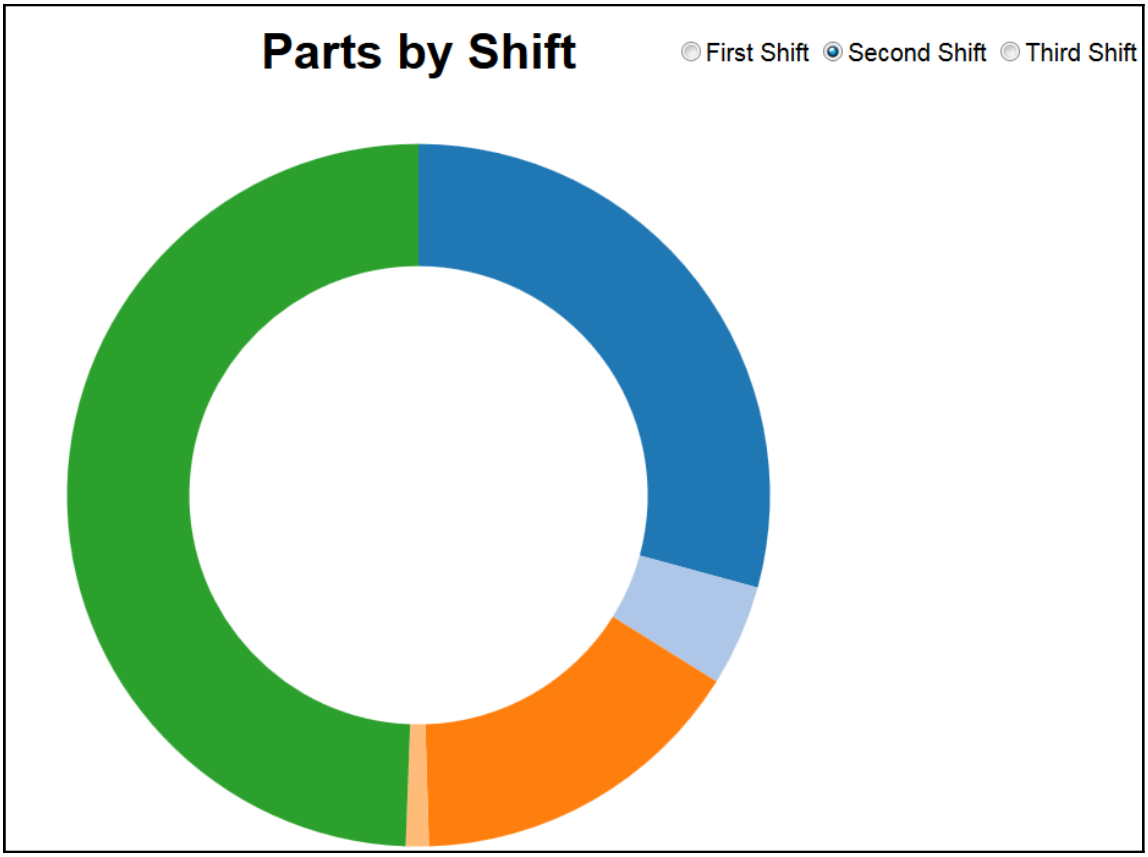
data.tsv

```
apples   oranges
53245    200
28479
19697    200
24037
40245    200
```



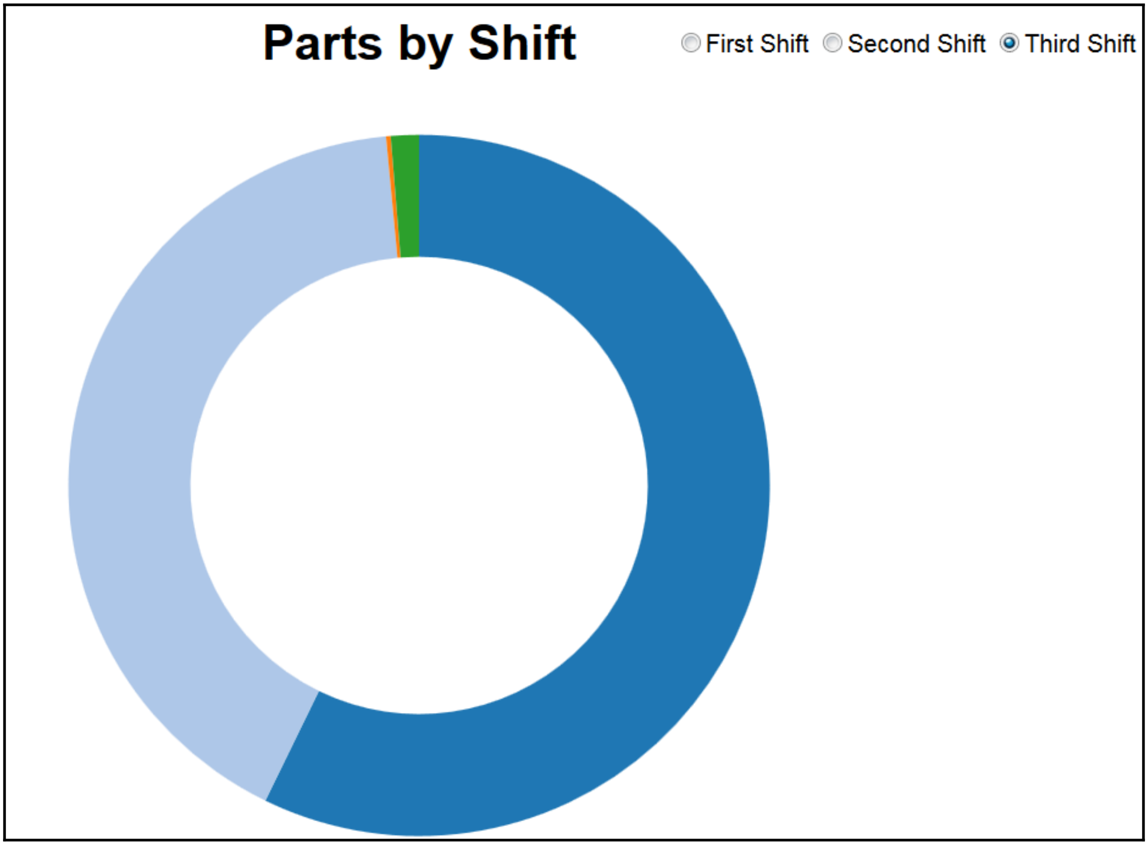
```
first    second   third
53245    53245    53245
28479    8479     38400
19697    28479    200
24037    1920
40245    90000    1200
```


Parts by Shift
● First Shift
○ Second Shift
○ Third Shift



Parts by Shift

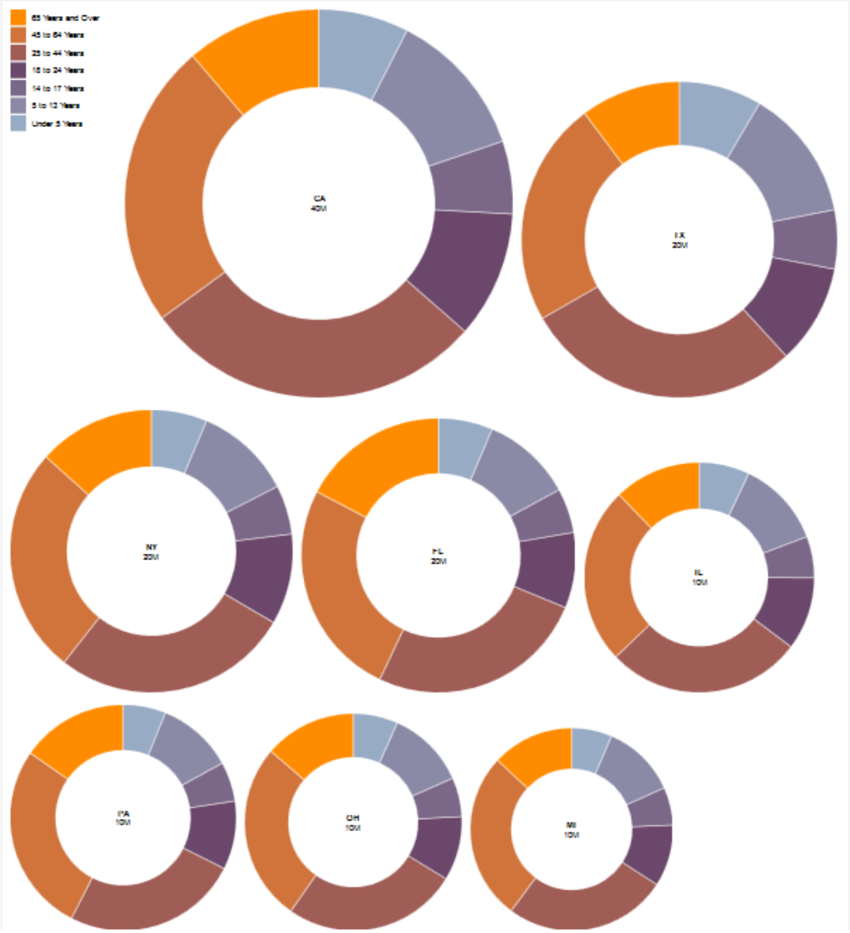
○ First Shift ● Second Shift ○ Third Shift

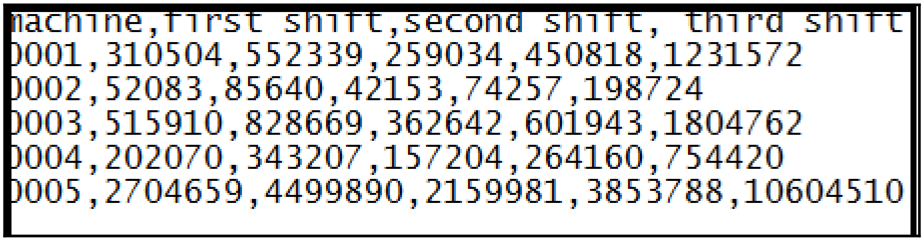


Parts by Shift

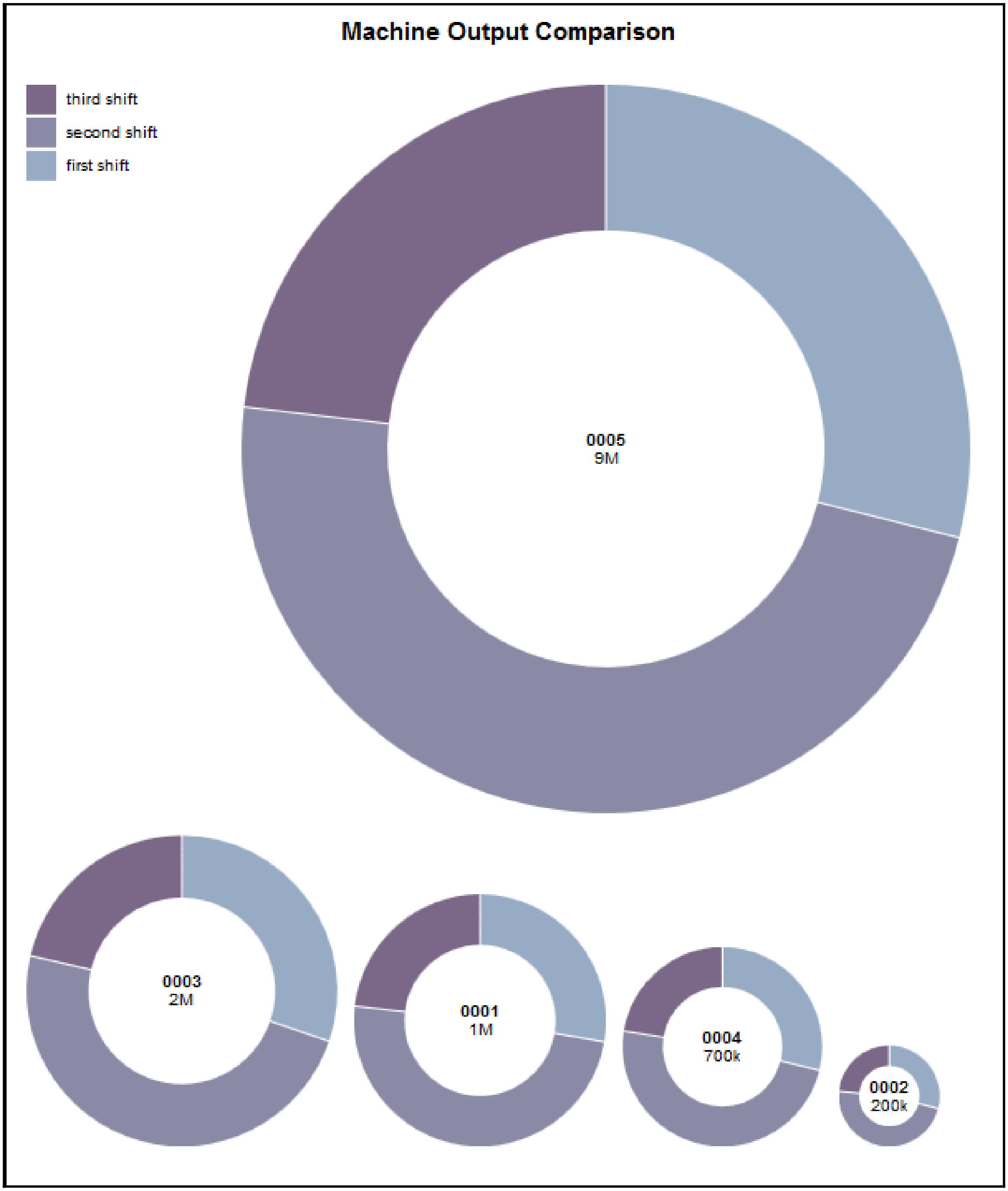
○ First Shift ○ Second Shift ● Third Shift

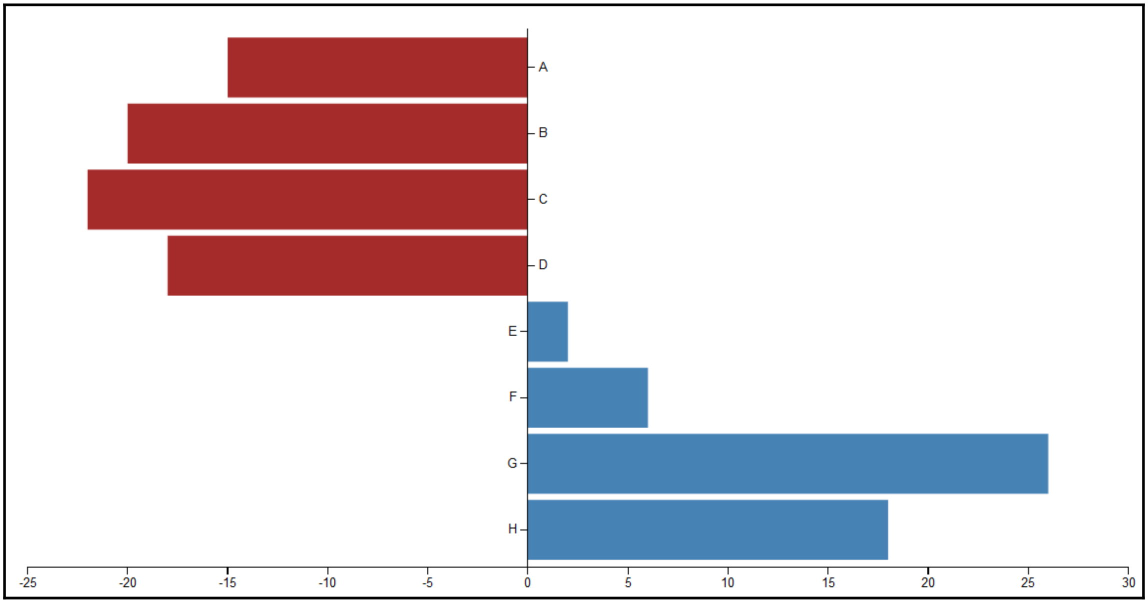
Sized Donut Multiples

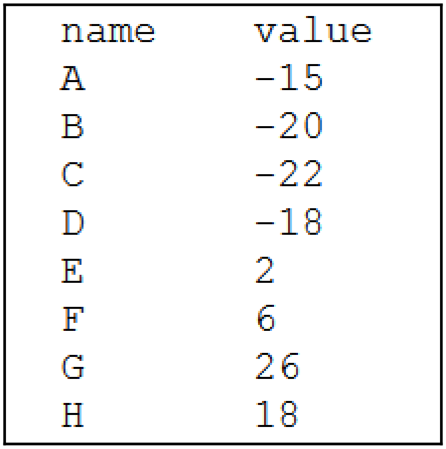




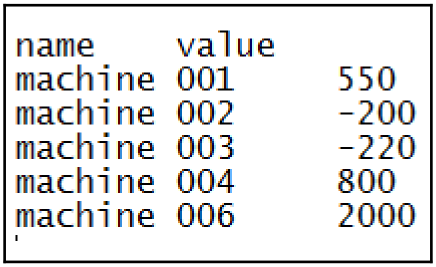
```
machine,first shift,second shift, third shift
0001,310504,552339,259034,450818,1231572
0002,52083,85640,42153,74257,198724
0003,515910,828669,362642,601943,1804762
0004,202070,343207,157204,264160,754420
0005,2704659,4499890,2159981,3853788,10604510
```


Machine Output Comparison
third shift
second shift
first shift
0005
9M
0003
2M
0001
1M
0004
700k
0002
200k

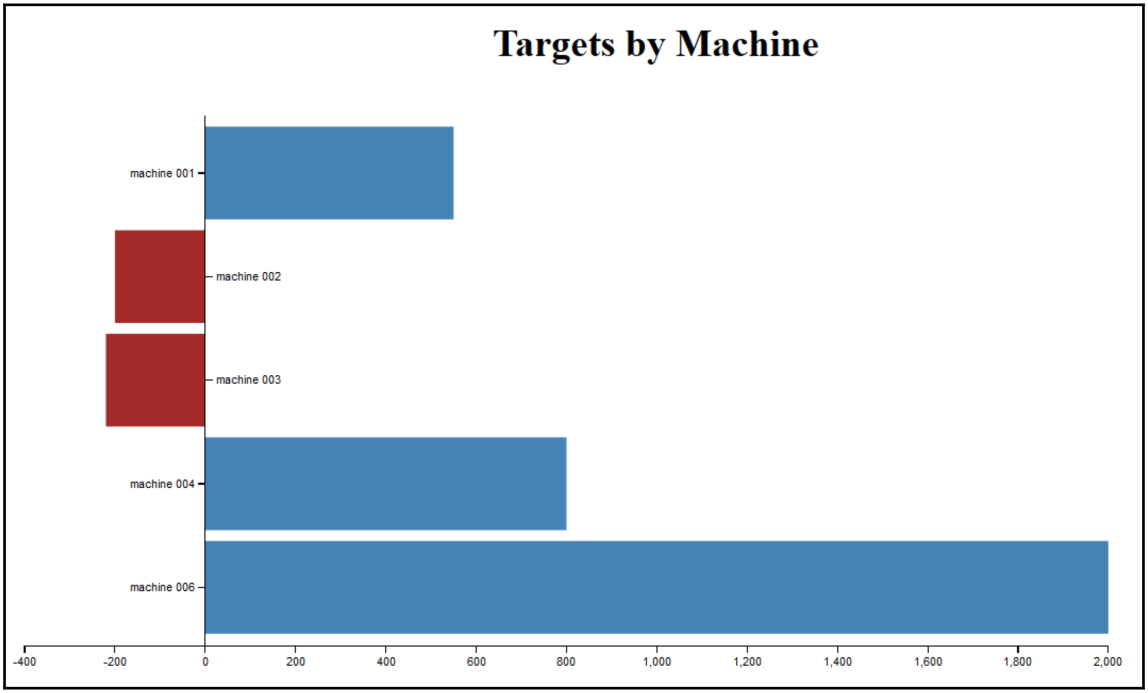


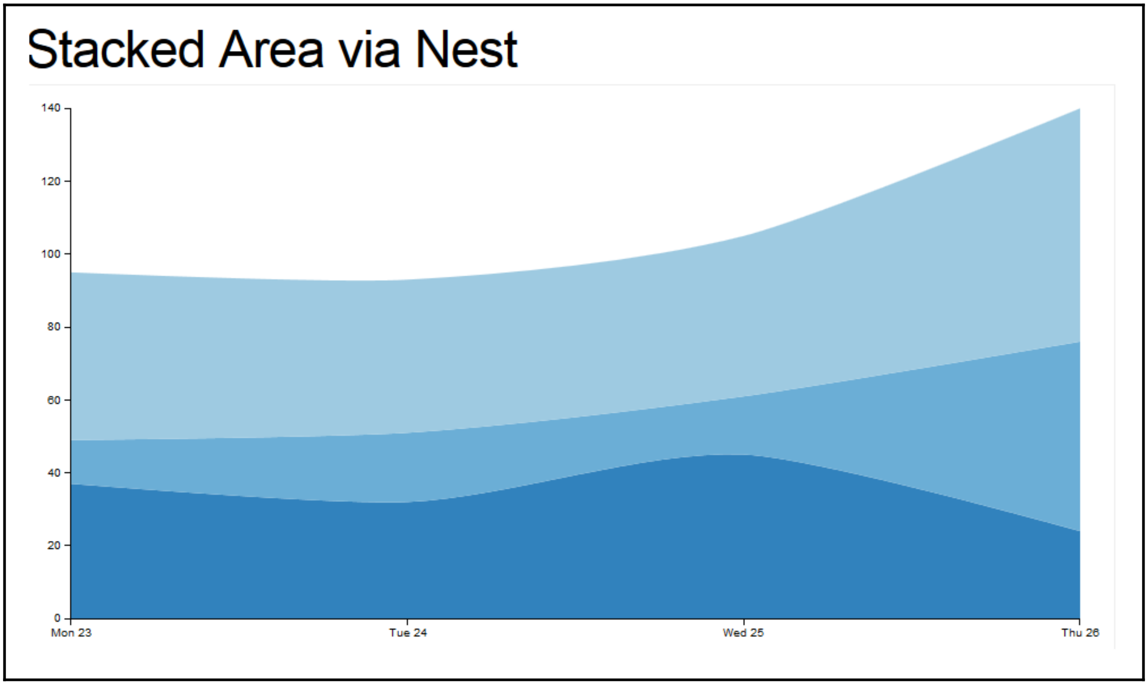


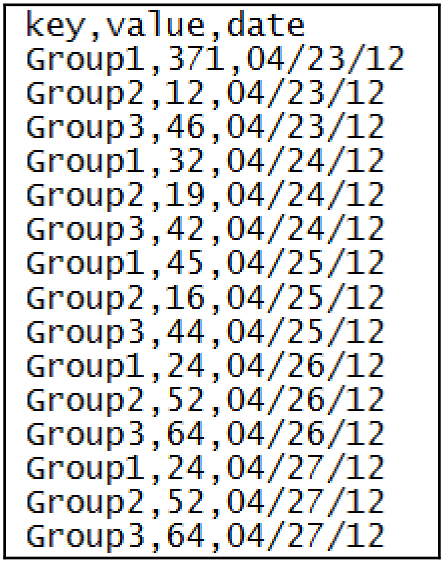
```
name      value
A         -15
B         -20
C         -22
D         -18
E         2
F         6
G         26
H         18
```



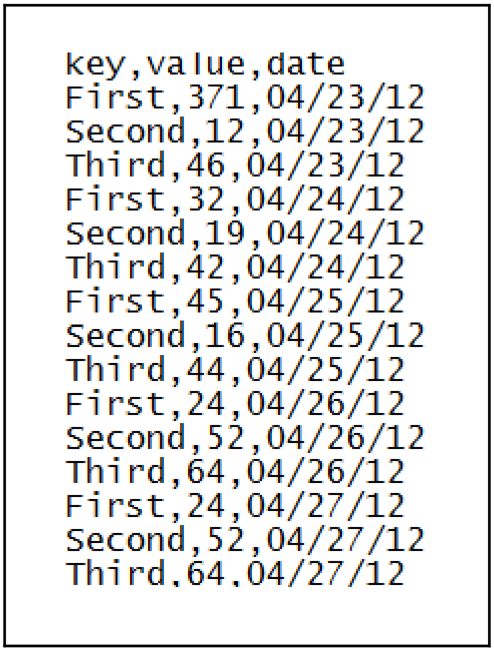
```
name    value
machine 001     550
machine 002     -200
machine 003     -220
machine 004     800
machine 006     2000
```


Targets by Machine
machine 001
machine 002
machine 003
machine 004
machine 006
-400
-200
0
200
400
600
800
1,000
1,200
1,400
1,600
1,800
2,000

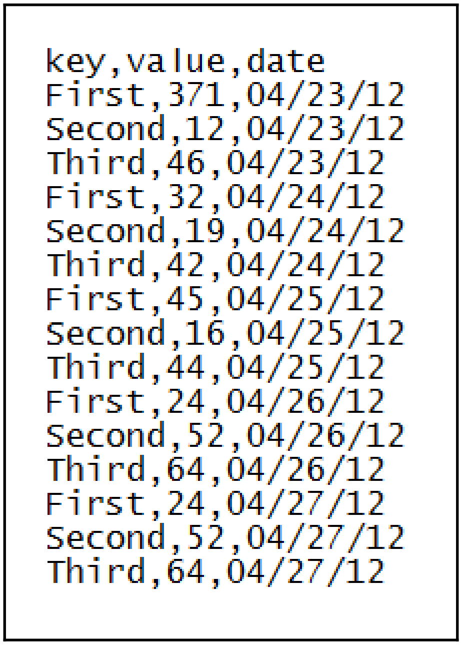

Stacked Area via Nest
140
120
100
80
60
40
20
0
Mon 23
Tue 24
Wed 25
Thu 26



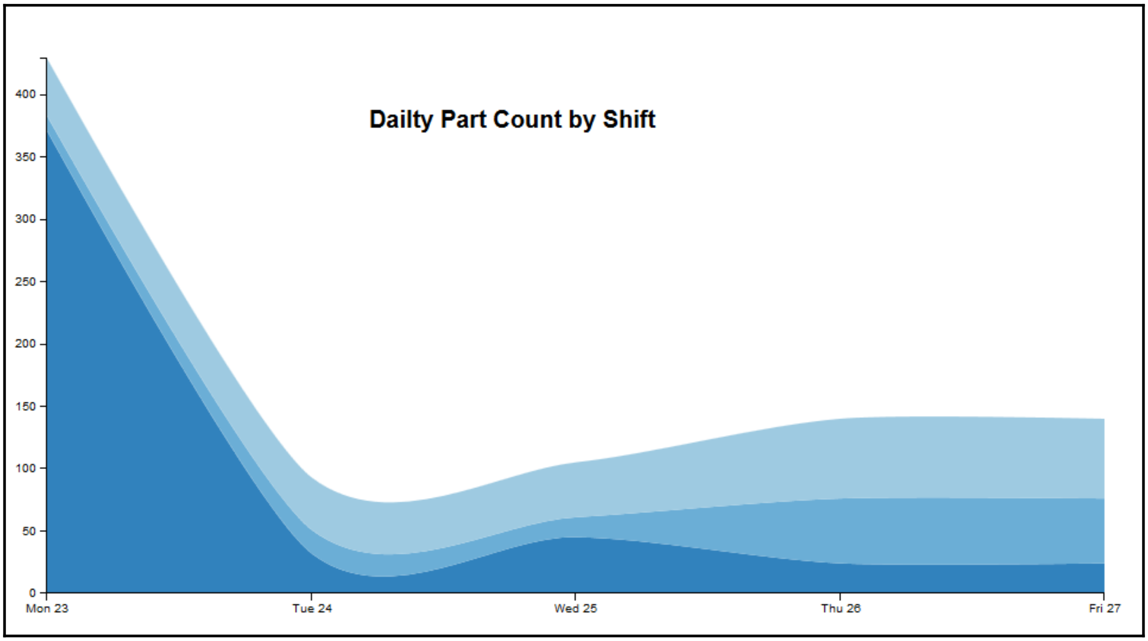
```
key,value,date
Group1,371,04/23/12
Group2,12,04/23/12
Group3,46,04/23/12
Group1,32,04/24/12
Group2,19,04/24/12
Group3,42,04/24/12
Group1,45,04/25/12
Group2,16,04/25/12
Group3,44,04/25/12
Group1,24,04/26/12
Group2,52,04/26/12
Group3,64,04/26/12
Group1,24,04/27/12
Group2,52,04/27/12
Group3,64,04/27/12
```



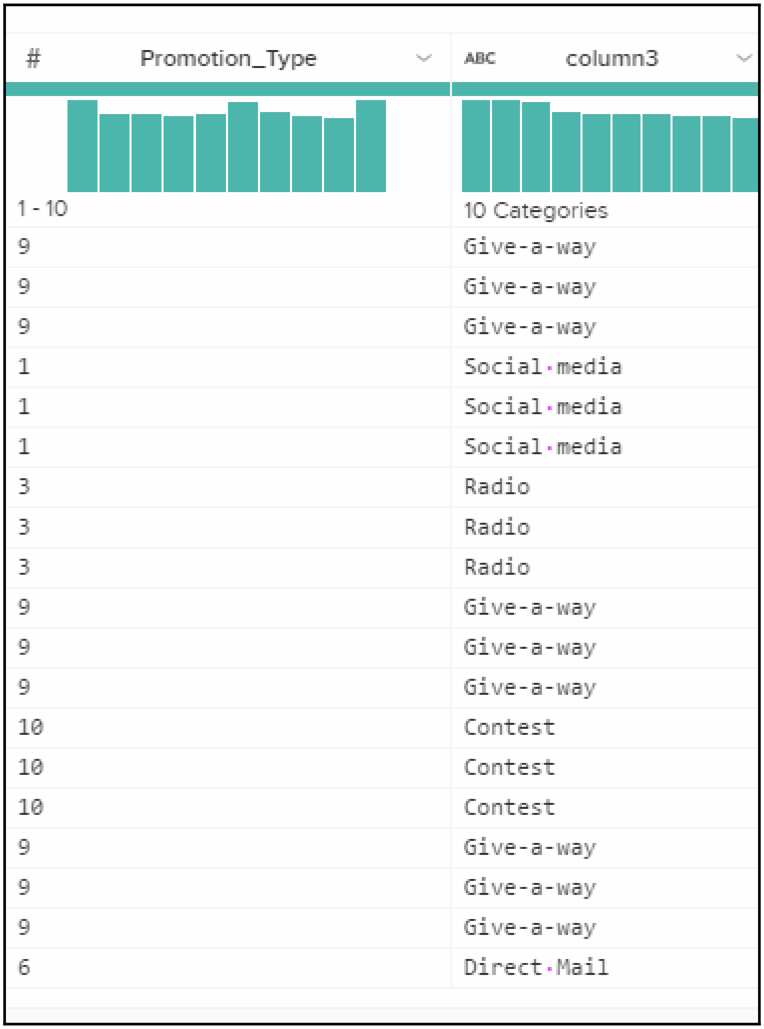
```
key,value,date
First,371,04/23/12
Second,12,04/23/12
Third,46,04/23/12
First,32,04/24/12
Second,19,04/24/12
Third,42,04/24/12
First,45,04/25/12
Second,16,04/25/12
Third,44,04/25/12
First,24,04/26/12
Second,52,04/26/12
Third,64,04/26/12
First,24,04/27/12
Second,52,04/27/12
Third.64.04/27/12
```



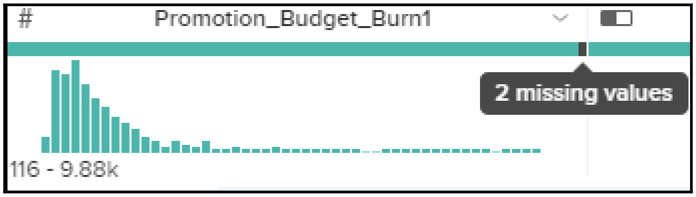
```
key,value,date
First,371,04/23/12
Second,12,04/23/12
Third,46,04/23/12
First,32,04/24/12
Second,19,04/24/12
Third,42,04/24/12
First,45,04/25/12
Second,16,04/25/12
Third,44,04/25/12
First,24,04/26/12
Second,52,04/26/12
Third,64,04/26/12
First,24,04/27/12
Second,52,04/27/12
Third,64,04/27/12
```

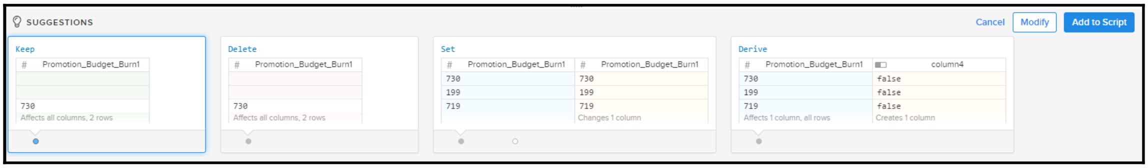


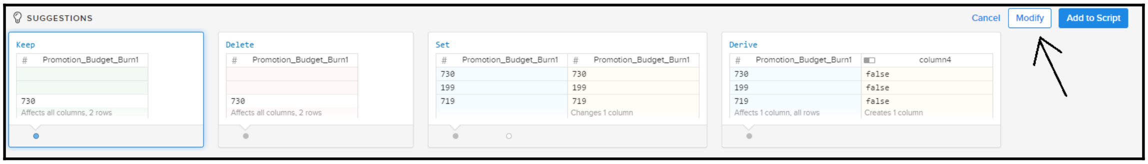
Chapter 6: Dashboards for Big Data – Tableau

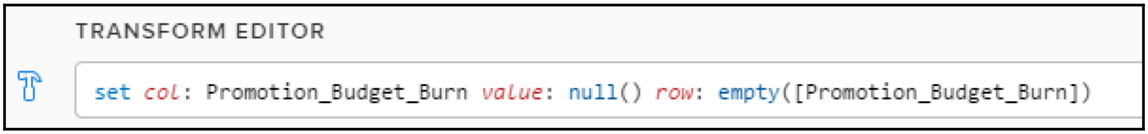


# Promotion_Type	ABC column3
1 - 10	10 Categories
9	Give-a-way
9	Give-a-way
9	Give-a-way
1	Social·media
1	Social·media
1	Social·media
3	Radio
3	Radio
3	Radio
9	Give-a-way
9	Give-a-way
9	Give-a-way
10	Contest
10	Contest
10	Contest
9	Give-a-way
9	Give-a-way
9	Give-a-way
6	Direct·Mail

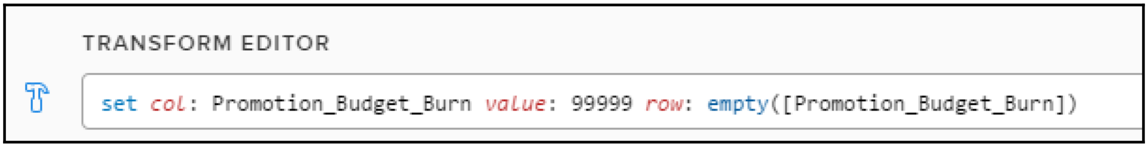




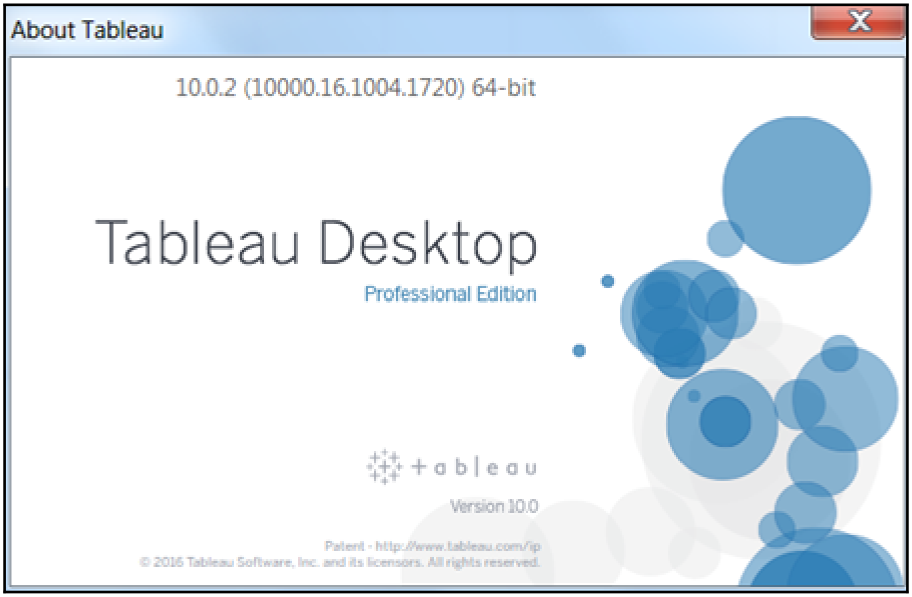




```
set col: Promotion_Budget_Burn value: null() row: empty([Promotion_Budget_Burn])
```



```
set col: Promotion_Budget_Burn value: 99999 row: empty([Promotion_Budget_Burn])
```



About Tableau

10.0.2 (10000.16.1004.1720) 64-bit

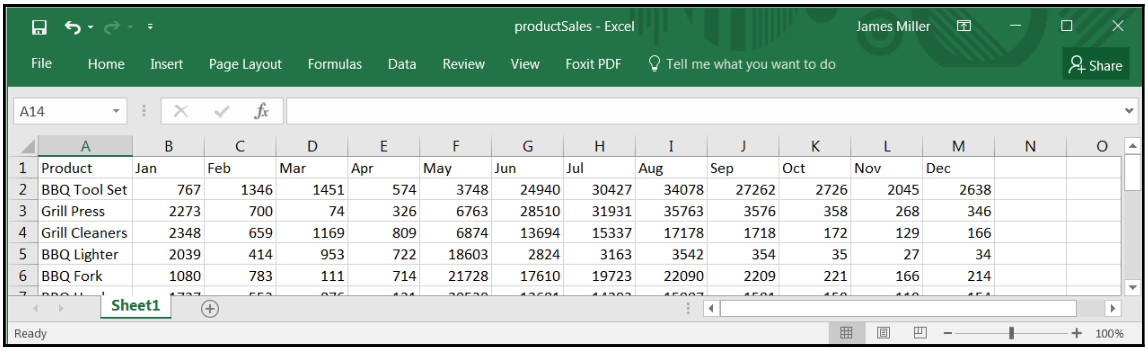
Tableau Desktop

Professional Edition

Version 10.0

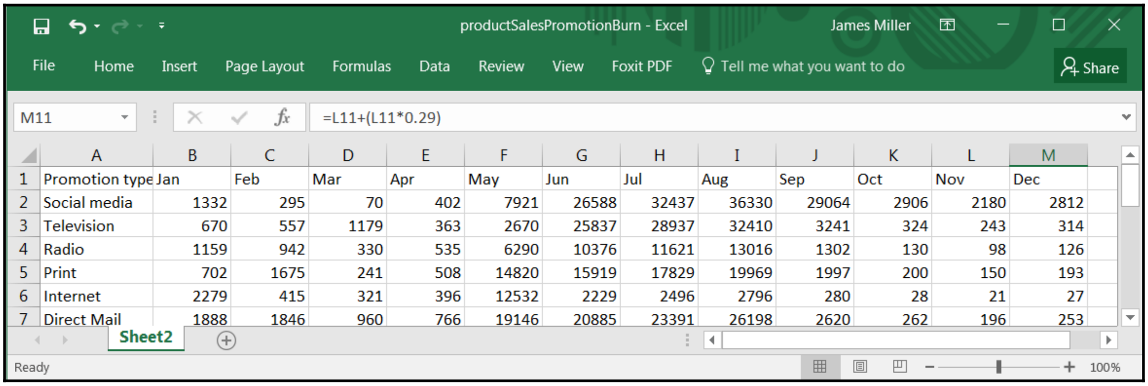
Patent - http://www.tableau.com/ip

© 2016 Tableau Software, Inc. and its licensors. All rights reserved.



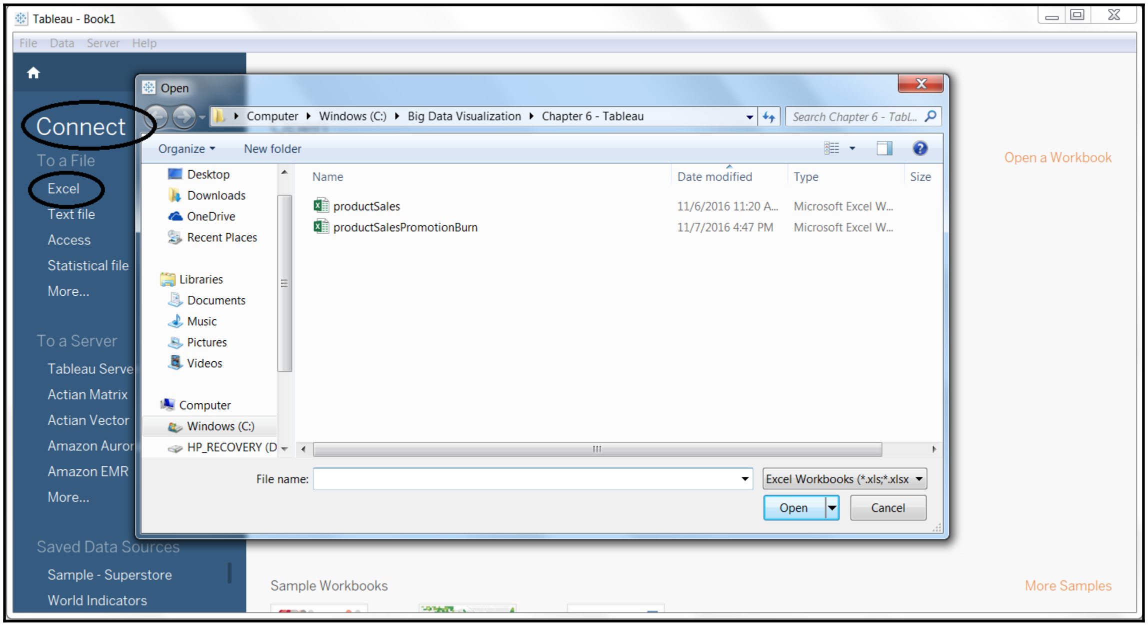
A14

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Product	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		
2	BBQ Tool Set	767	1346	1451	574	3748	24940	30427	34078	27262	2726	2045	2638		
3	Grill Press	2273	700	74	326	6763	28510	31931	35763	3576	358	268	346		
4	Grill Cleaners	2348	659	1169	809	6874	13694	15337	17178	1718	172	129	166		
5	BBQ Lighter	2039	414	953	722	18603	2824	3163	3542	354	35	27	34		
6	BBQ Fork	1080	783	111	714	21728	17610	19723	22090	2209	221	166	214		



M11 =L11+(L11*0.29)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Promotion type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2	Social media	1332	295	70	402	7921	26588	32437	36330	29064	2906	2180	2812
3	Television	670	557	1179	363	2670	25837	28937	32410	3241	324	243	314
4	Radio	1159	942	330	535	6290	10376	11621	13016	1302	130	98	126
5	Print	702	1675	241	508	14820	15919	17829	19969	1997	200	150	193
6	Internet	2279	415	321	396	12532	2229	2496	2796	280	28	21	27
7	Direct Mail	1888	1846	960	766	19146	20885	23391	26198	2620	262	196	253



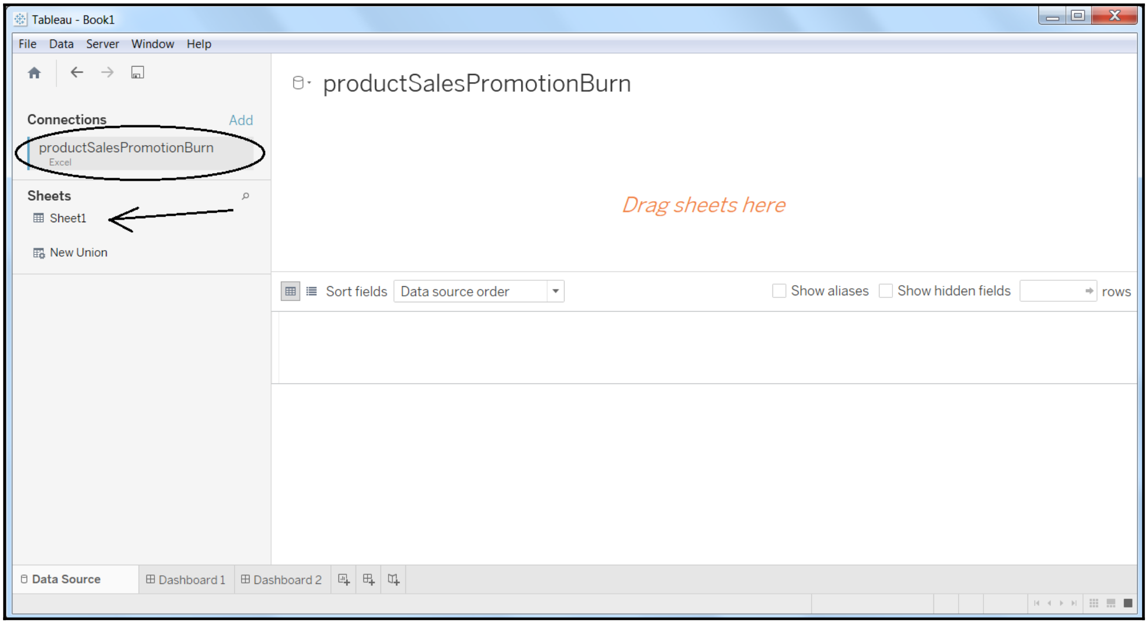
Computer ▸ Windows (C:) ▸ Big Data Visualization ▸ Chapter 6 - Tableau

Search Chapter 6 - Tabl...

Name	Date modified	Type	Size
productSales	11/6/2016 11:20 A...	Microsoft Excel W...	
productSalesPromotionBurn	11/7/2016 4:47 PM	Microsoft Excel W...	

File name:

Excel Workbooks (*.xls;*.xlsx

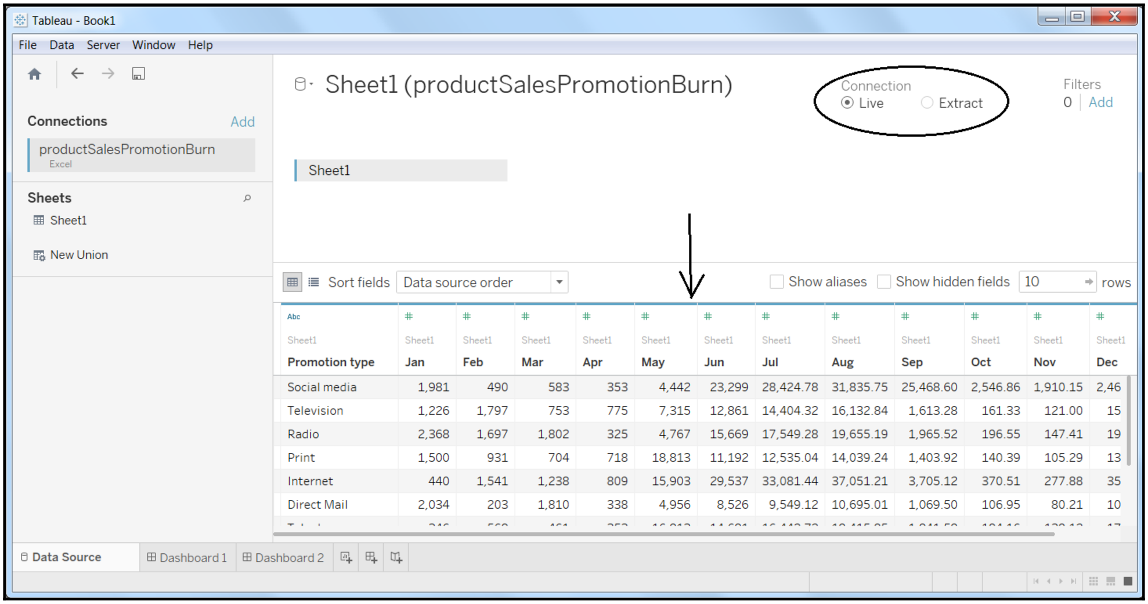


productSalesPromotionBurn

Connections: productSalesPromotionBurn (Excel)

Sheets: Sheet1

Sort fields: Data source order



Sheet1 (productSalesPromotionBurn)

Connection ● Live ○ Extract

Sort fields: Data source order □ Show aliases □ Show hidden fields 10 rows

Promotion type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Social media	1,981	490	583	353	4,442	23,299	28,424.78	31,835.75	25,468.60	2,546.86	1,910.15	2,46
Television	1,226	1,797	753	775	7,315	12,861	14,404.32	16,132.84	1,613.28	161.33	121.00	15
Radio	2,368	1,697	1,802	325	4,767	15,669	17,549.28	19,655.19	1,965.52	196.55	147.41	19
Print	1,500	931	704	718	18,813	11,192	12,535.04	14,039.24	1,403.92	140.39	105.29	13
Internet	440	1,541	1,238	809	15,903	29,537	33,081.44	37,051.21	3,705.12	370.51	277.88	35
Direct Mail	2,034	203	1,810	338	4,956	8,526	9,549.12	10,695.01	1,069.50	106.95	80.21	10

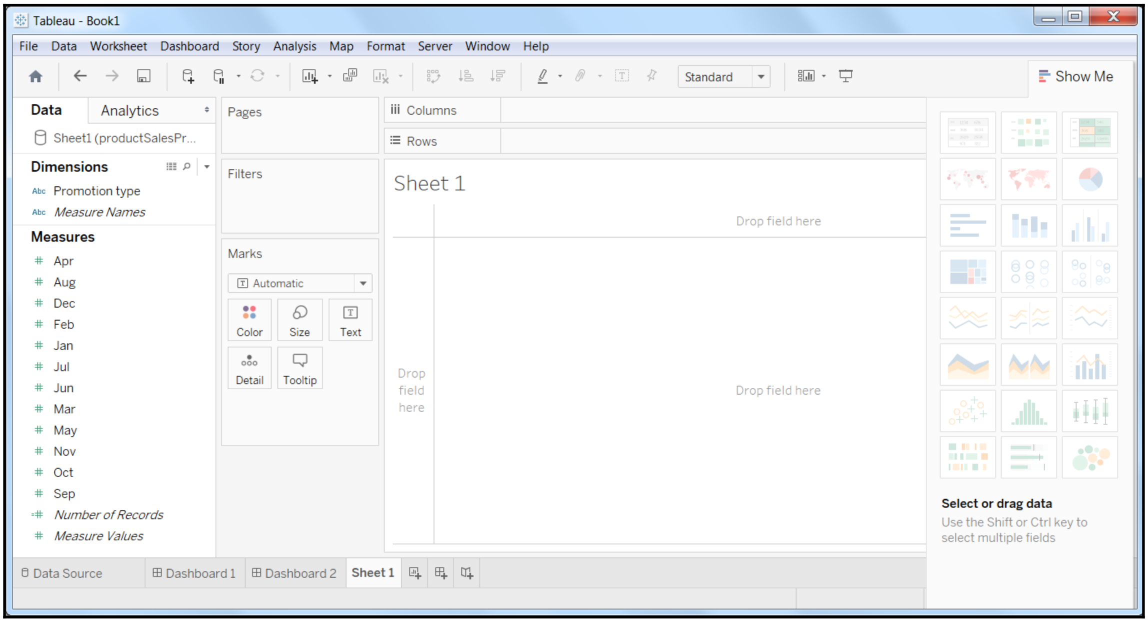
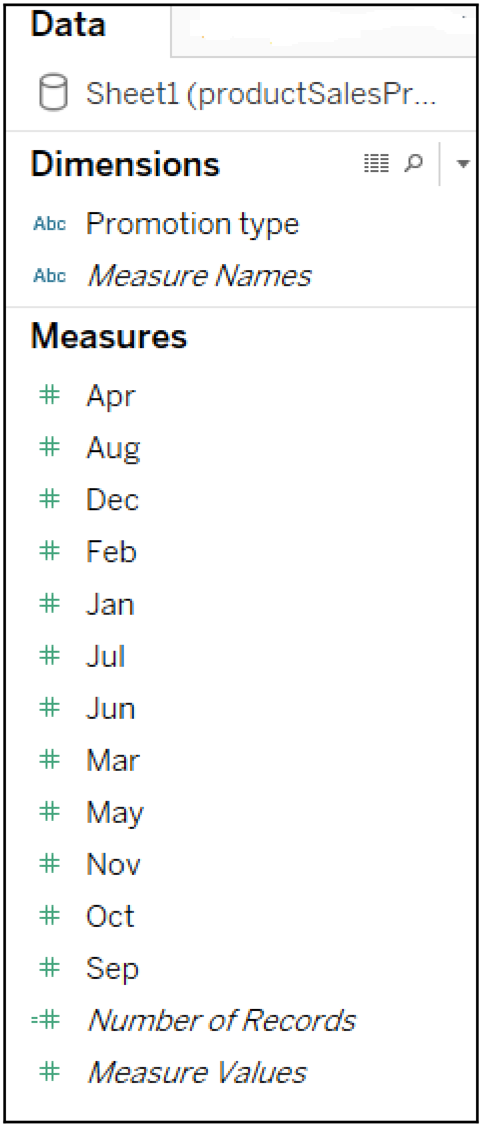
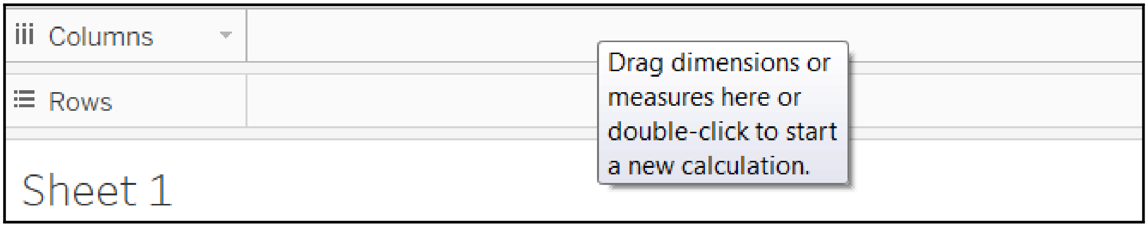


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
Sheet1 (productSalesPr...
Pages
Columns
Rows
Filters
Marks
Automatic
Color Size Text
Detail Tooltip
Sheet 1
Drop field here
Drop field here
Drop field here
Select or drag data
Use the Shift or Ctrl key to select multiple fields
Data Source Dashboard 1 Dashboard 2 Sheet 1

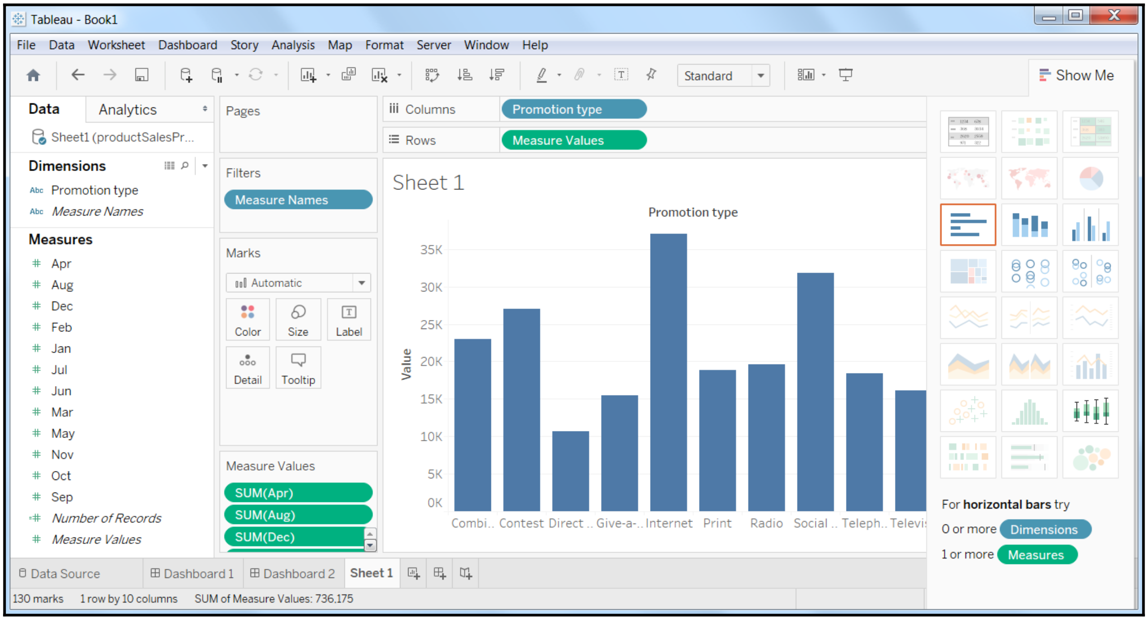
Dimensions
- Promotion type
- *Measure Names*

Measures
- Apr
- Aug
- Dec
- Feb
- Jan
- Jul
- Jun
- Mar
- May
- Nov
- Oct
- Sep
- *Number of Records*
- *Measure Values*

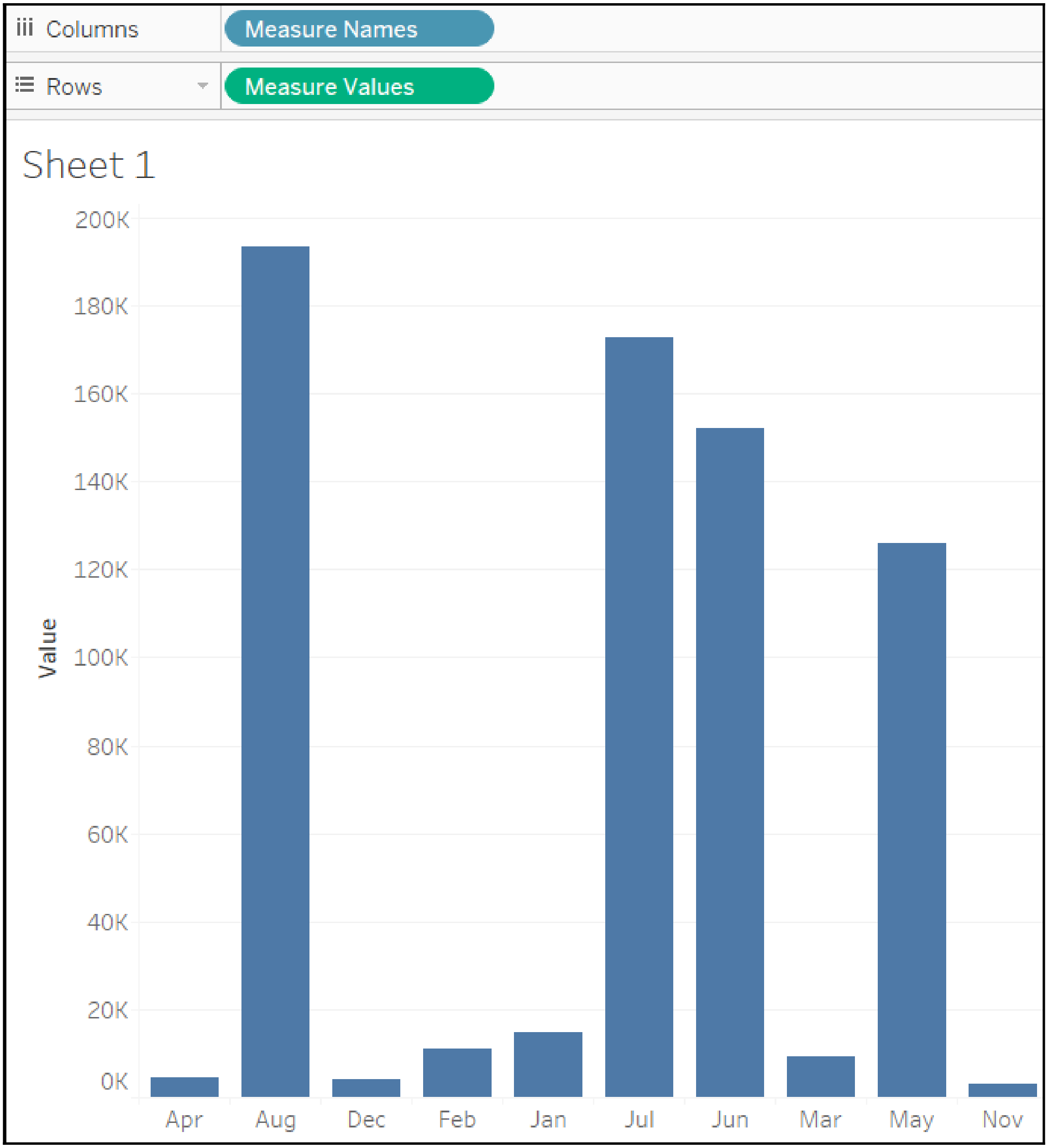




Drag dimensions or measures here or double-click to start a new calculation.



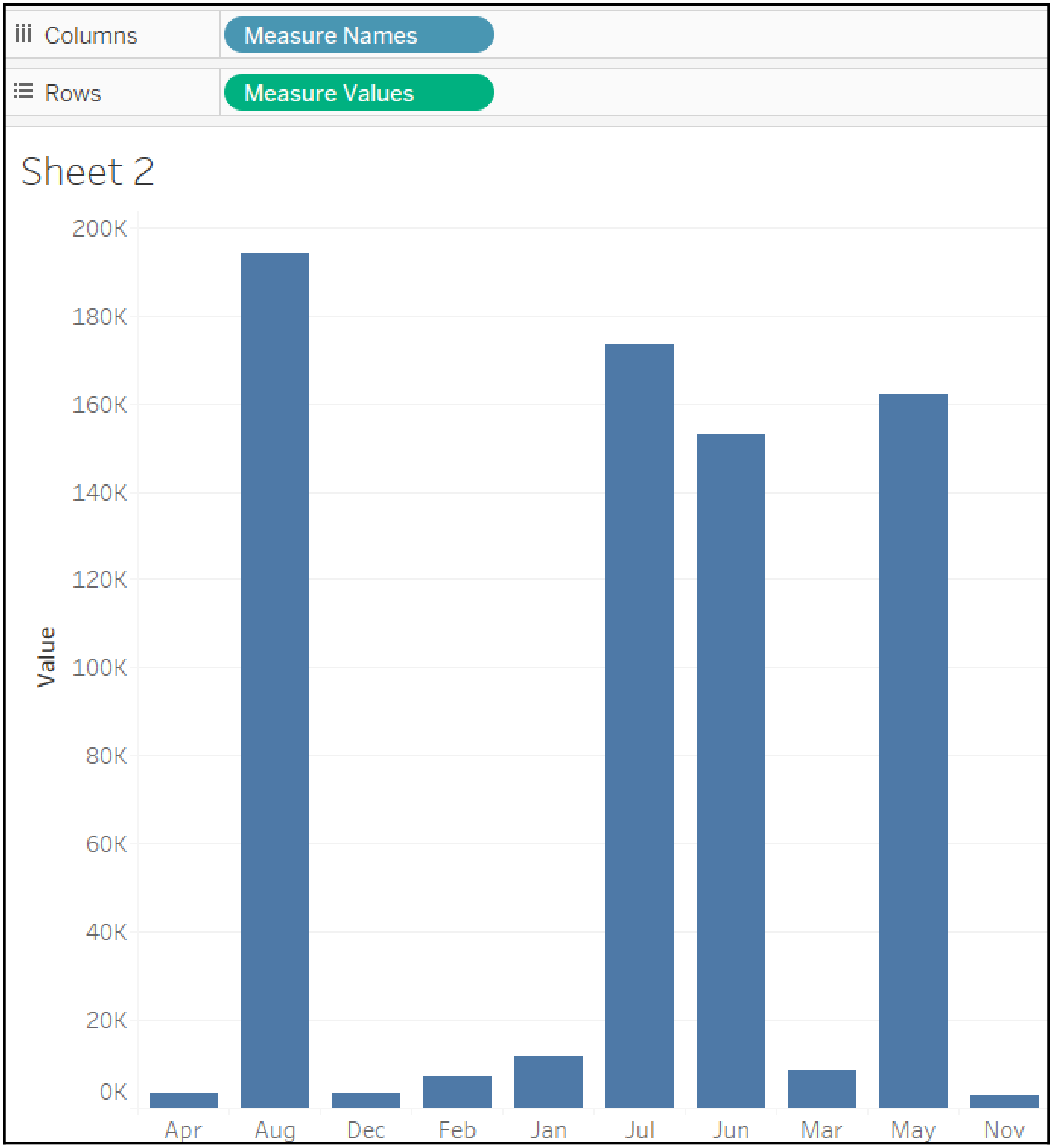
130 marks   1 row by 10 columns   SUM of Measure Values: 736.175



Columns: Measure Names

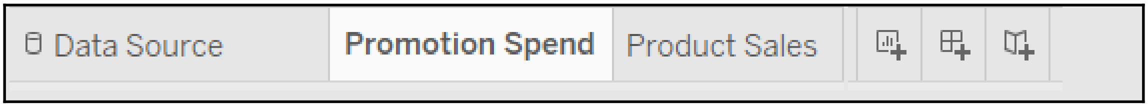
Rows: Measure Values

Sheet 1



Columns: Measure Names

Rows: Measure Values



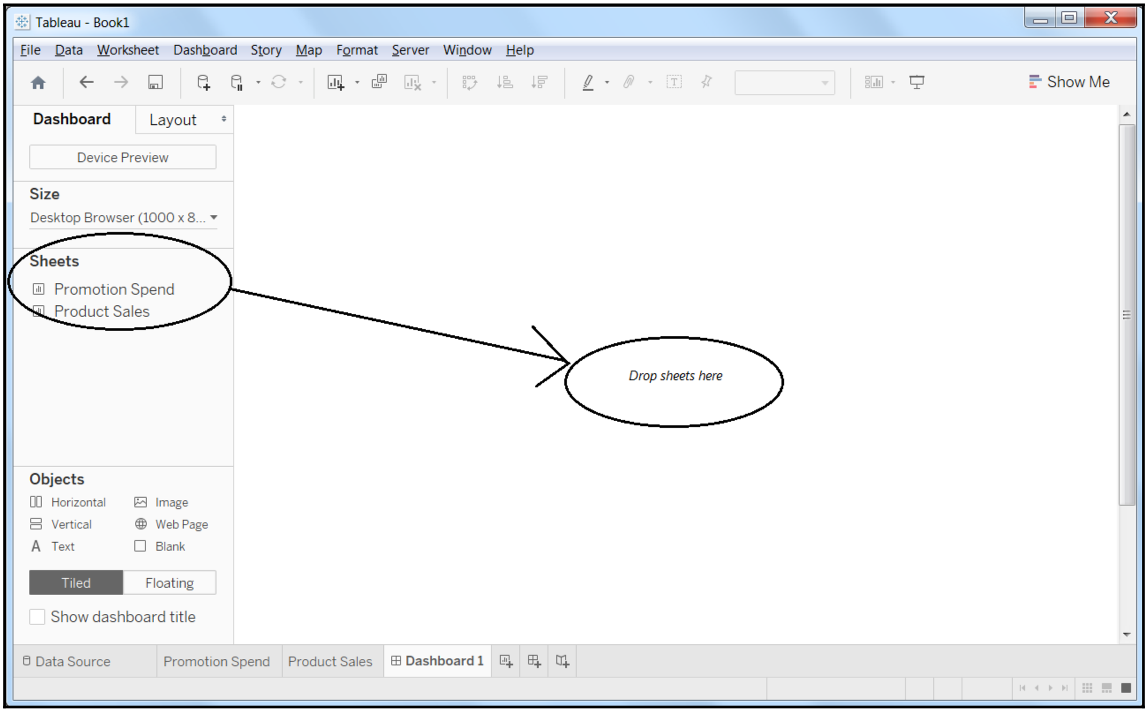


Tableau - Book1
File Data Worksheet Dashboard Story Map Format Server Window Help
Show Me
Dashboard
Layout
Device Preview
Size
Desktop Browser (1000 x 8...
Sheets
Promotion Spend
Product Sales
Drop sheets here
Objects
Horizontal
Image
Vertical
Web Page
Text
Blank
Tiled
Floating
Show dashboard title
Data Source
Promotion Spend
Product Sales
Dashboard 1

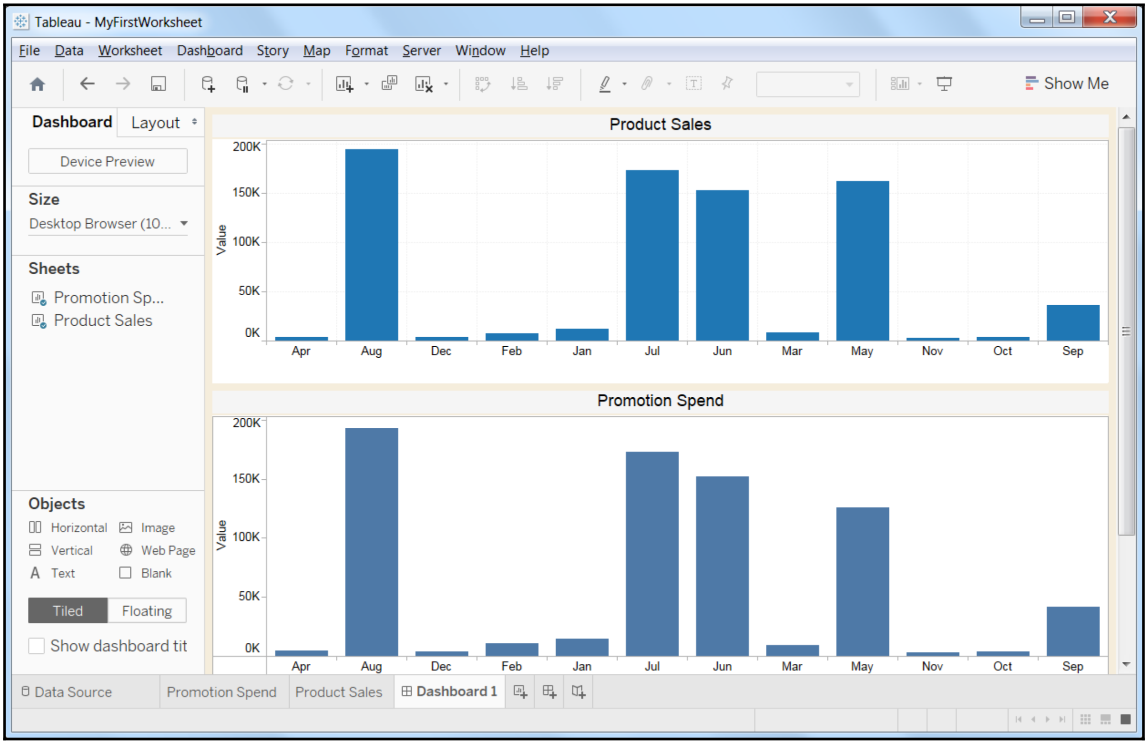
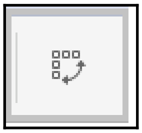
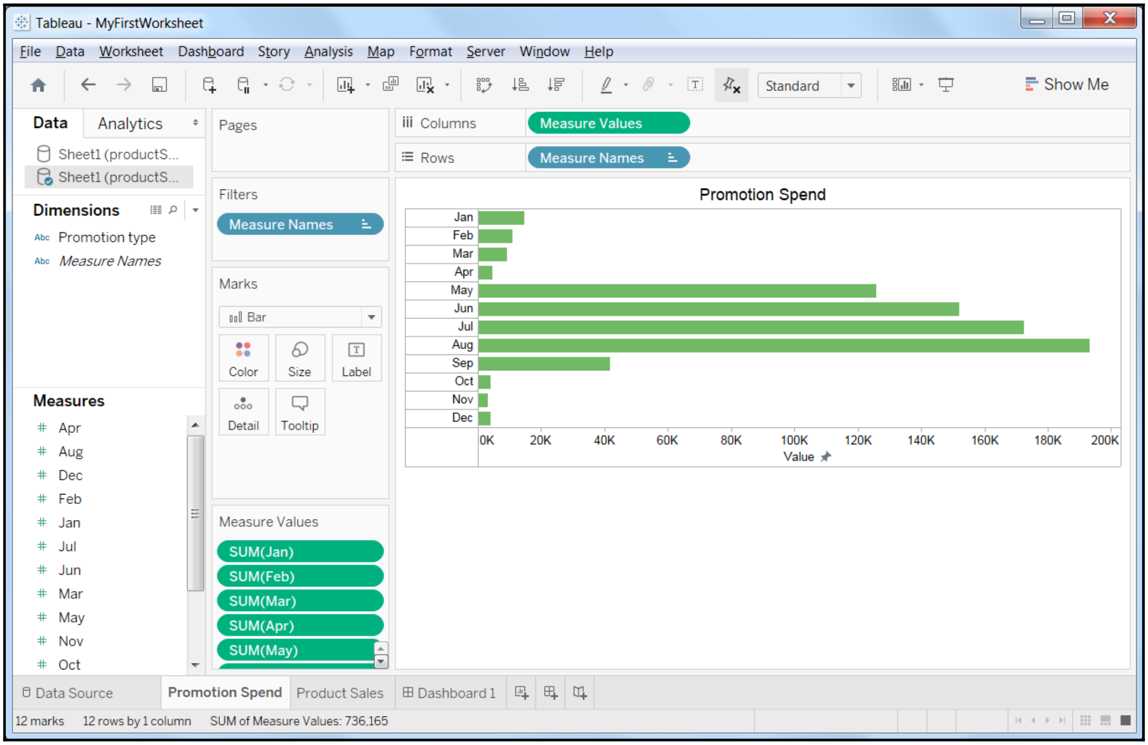
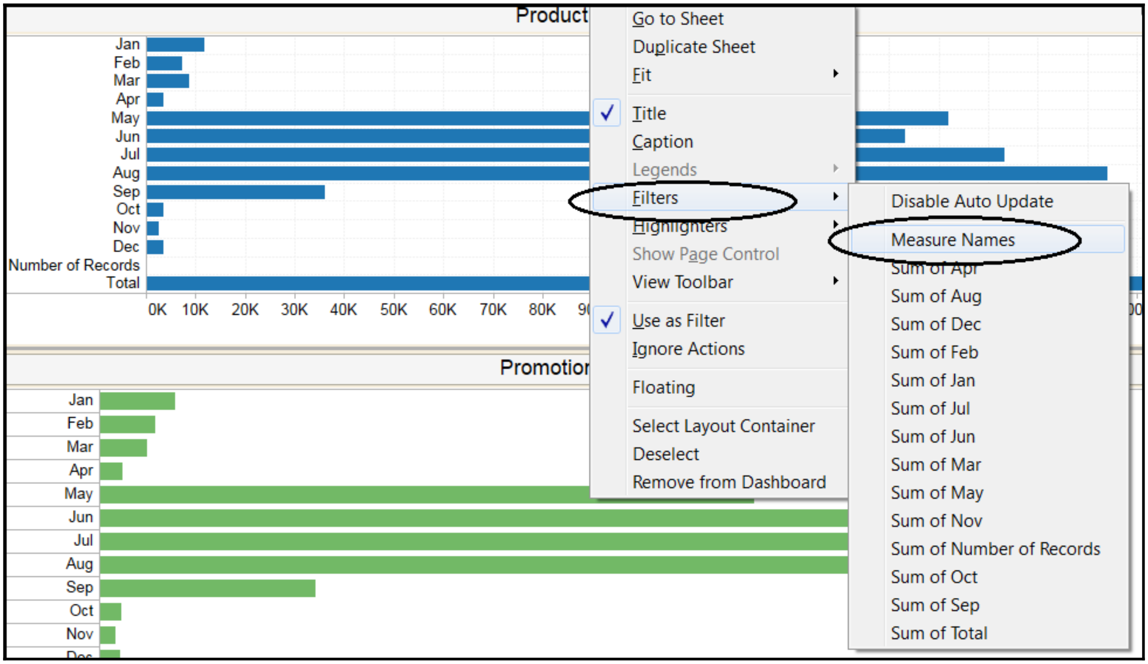


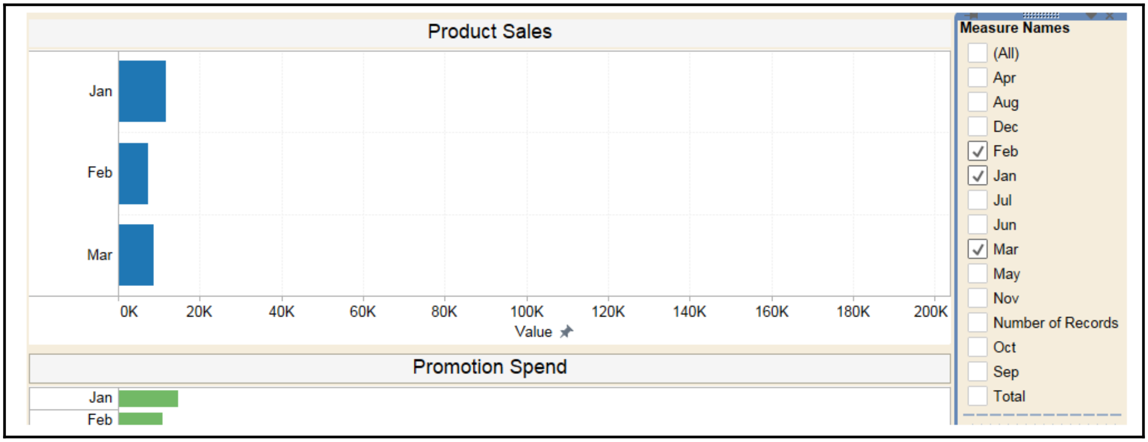
Tableau - MyFirstWorksheet
File Data Worksheet Dashboard Story Map Format Server Window Help
Show Me
Dashboard
Layout
Device Preview
Size
Desktop Browser (10...
Sheets
Promotion Sp...
Product Sales
Objects
Horizontal
Image
Vertical
Web Page
Text
Blank
Tiled
Floating
Show dashboard tit
Product Sales
Value
200K
150K
100K
50K
0K
Apr Aug Dec Feb Jan Jul Jun Mar May Nov Oct Sep
Promotion Spend
Value
200K
150K
100K
50K
0K
Apr Aug Dec Feb Jan Jul Jun Mar May Nov Oct Sep
Data Source
Promotion Spend
Product Sales
Dashboard 1





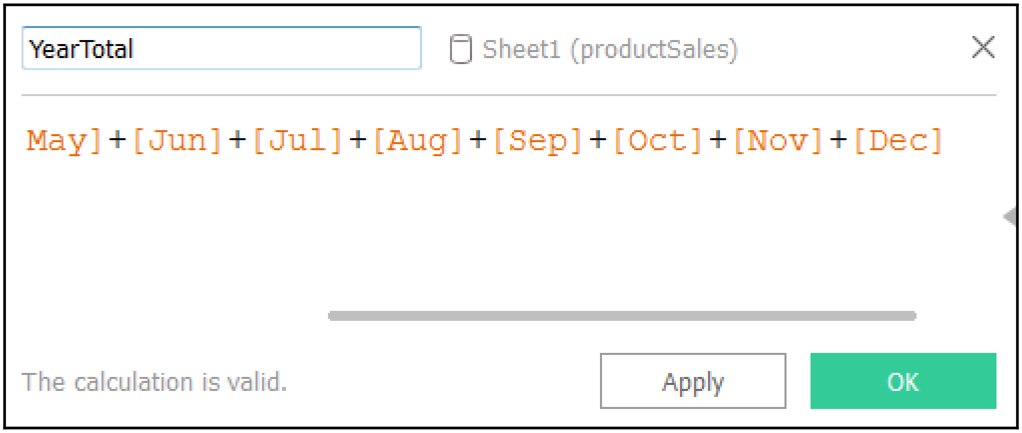
12 marks 12 rows by 1 column SUM of Measure Values: 736,165





Measure Names

- □ (All)
- □ Apr
- □ Aug
- □ Dec
- ☑ Feb
- ☑ Jan
- □ Jul
- □ Jun
- ☑ Mar
- □ May
- □ Nov
- □ Number of Records
- □ Oct
- □ Sep
- □ Total



YearTotal

Sheet1 (productSales)

```
May]+[Jun]+[Jul]+[Aug]+[Sep]+[Oct]+[Nov]+[Dec]
```

The calculation is valid.

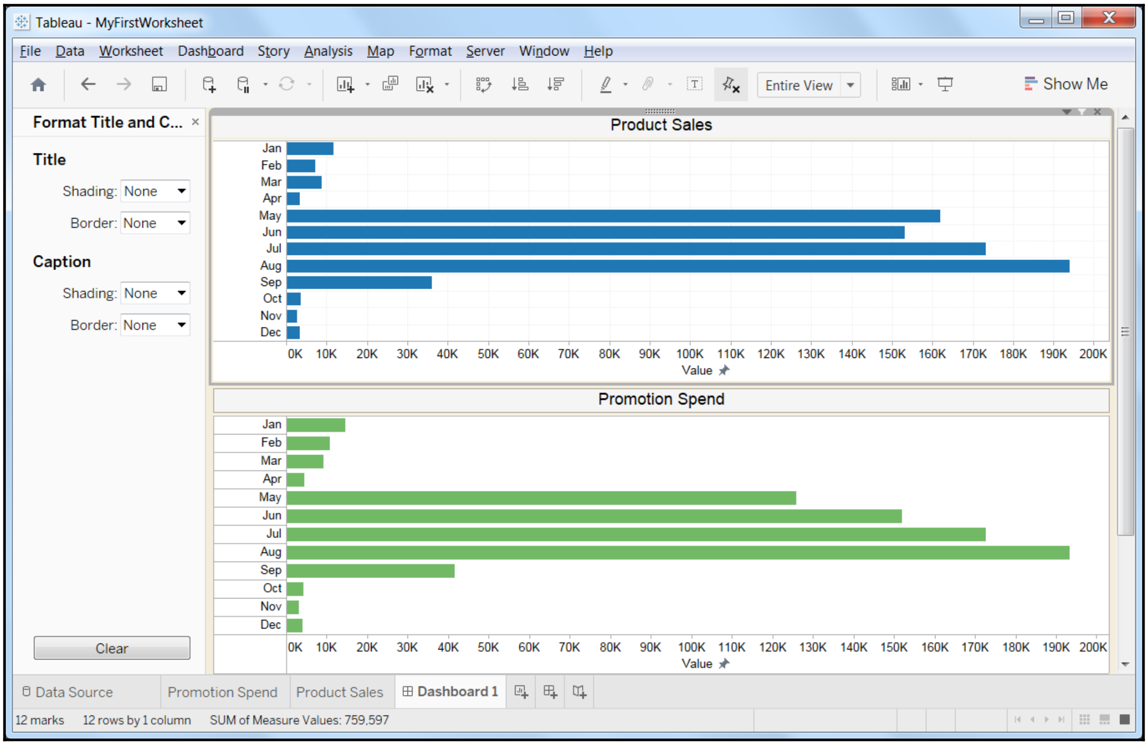


Tableau - MyFirstWorksheet

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Format Title and C...

Title

Shading: None

Border: None

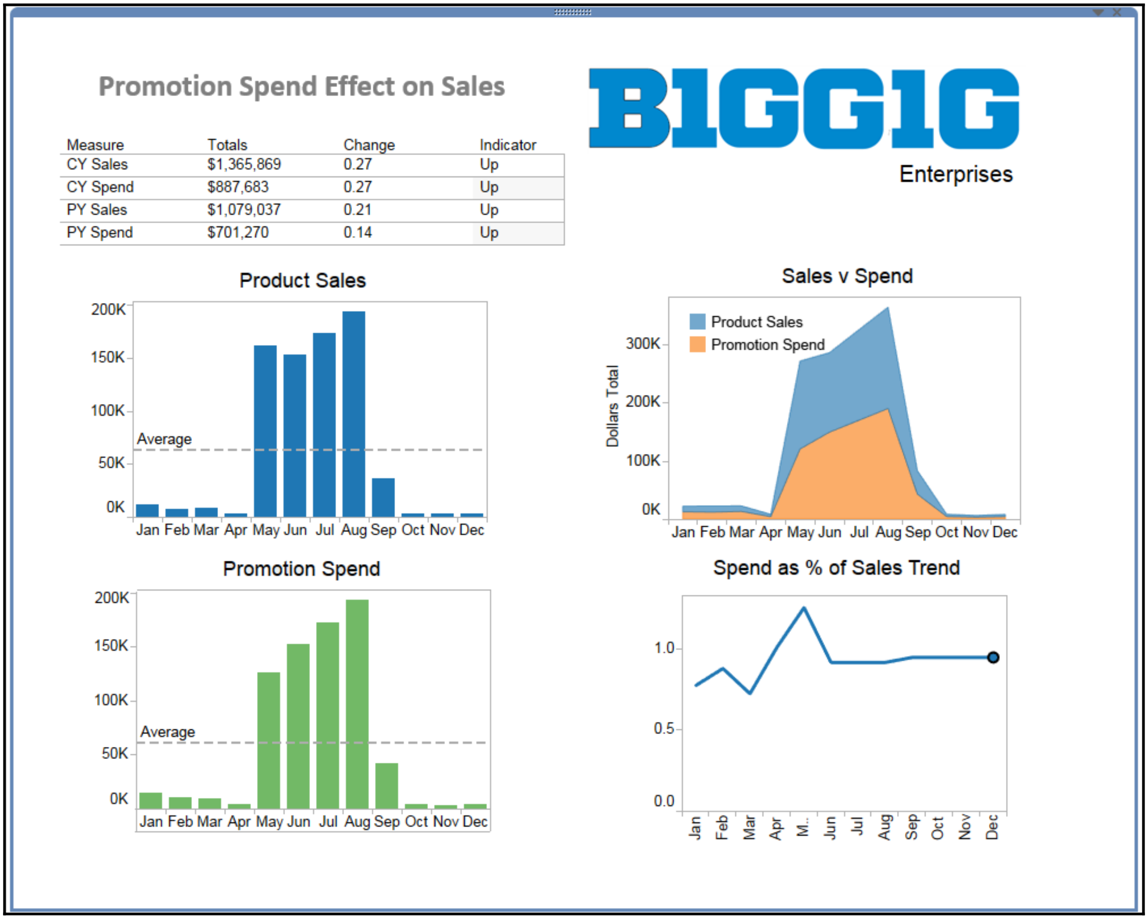
Caption

Shading: None

Border: None

Clear
Product Sales
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
0K 10K 20K 30K 40K 50K 60K 70K 80K 90K 100K 110K 120K 130K 140K 150K 160K 170K 180K 190K 200K
Value
Promotion Spend
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
0K 10K 20K 30K 40K 50K 60K 70K 80K 90K 100K 110K 120K 130K 140K 150K 160K 170K 180K 190K 200K
Value
Data Source
Promotion Spend
Product Sales
Dashboard 1

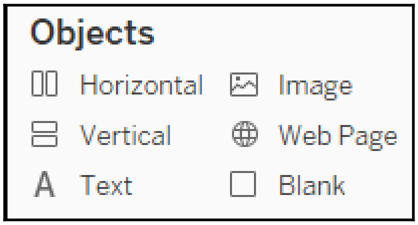
12 marks 12 rows by 1 column SUM of Measure Values: 759,597

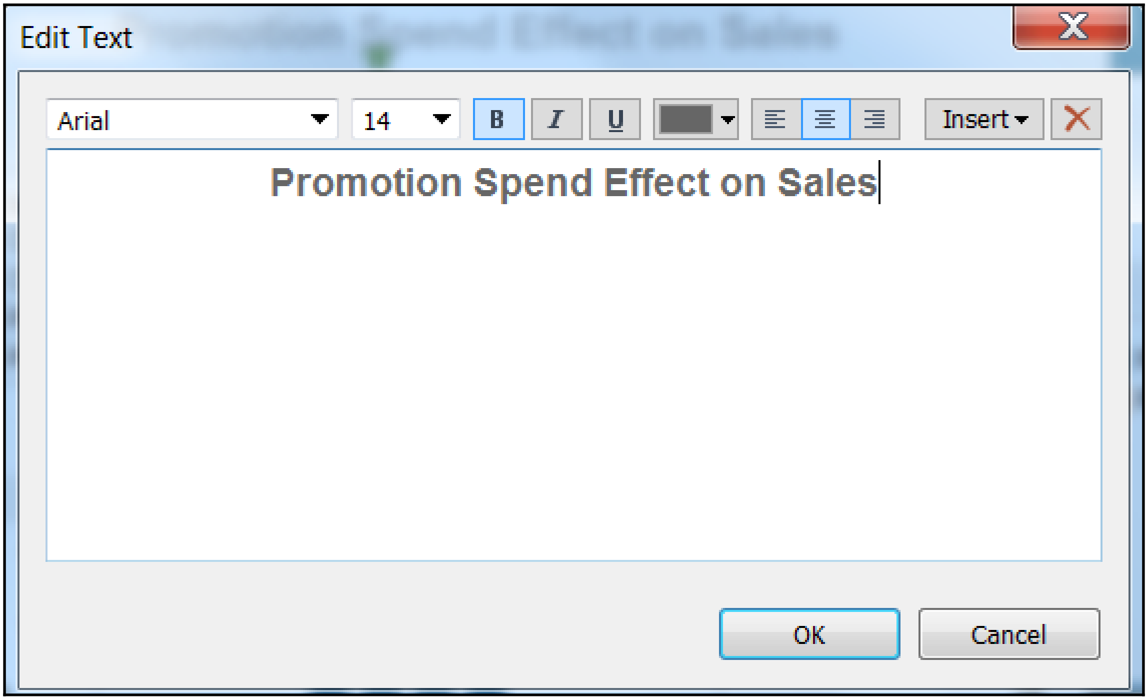


Promotion Spend Effect on Sales

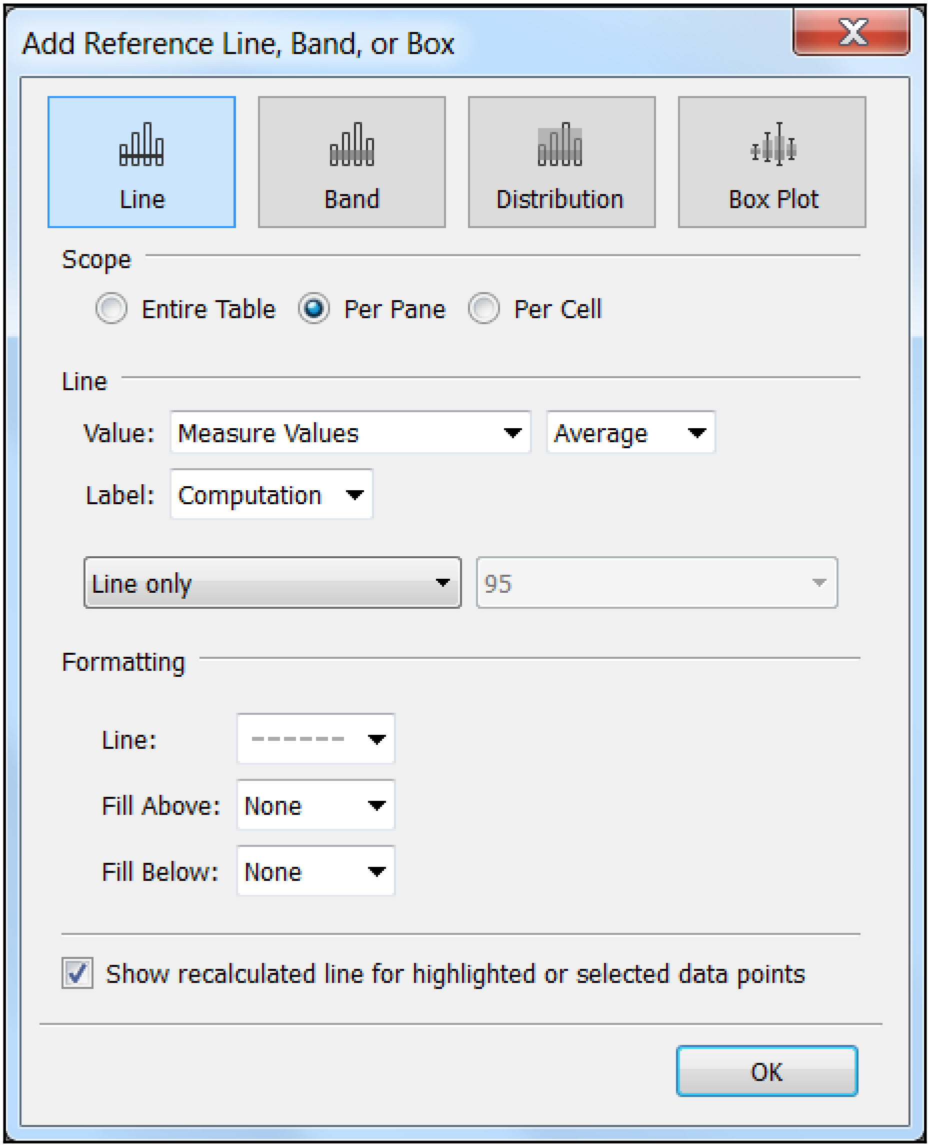
Measure	Totals	Change	Indicator
CY Sales	$1,365,869	0.27	Up
CY Spend	$887,683	0.27	Up
PY Sales	$1,079,037	0.21	Up
PY Spend	$701,270	0.14	Up

BIGGIG
Enterprises
Product Sales
200K
150K
100K
50K
0K
Average
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Sales v Spend
Product Sales
Promotion Spend
Dollars Total
300K
200K
100K
0K
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Promotion Spend
200K
150K
100K
50K
0K
Average
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Spend as % of Sales Trend
1.0
0.5
0.0
Jan Feb Mar Apr M.. Jun Jul Aug Sep Oct Nov Dec





Promotion Spend Effect on Sales



Add Reference Line, Band, or Box

Scope

○ Entire Table ● Per Pane ○ Per Cell

Line

Value: Measure Values ▾ Average ▾

Label: Computation ▾

Line only ▾ 95 ▾

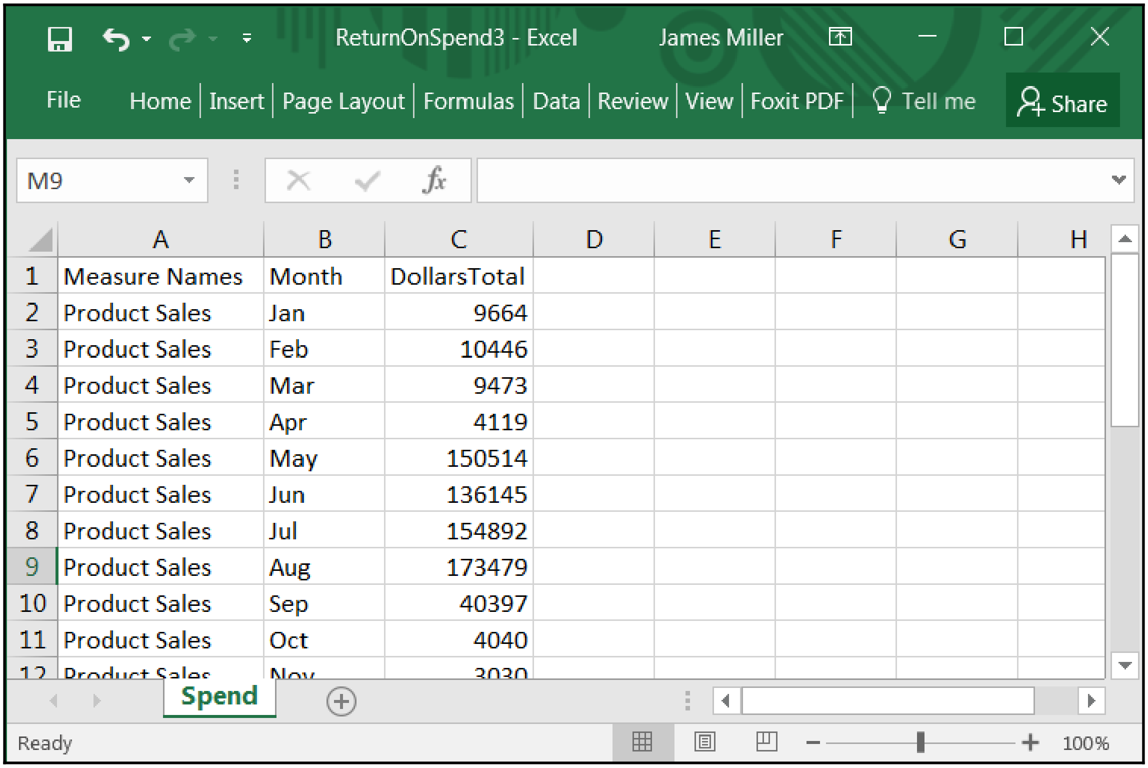
Formatting

Line: ------ ▾

Fill Above: None ▾

Fill Below: None ▾

☑ Show recalculated line for highlighted or selected data points



	A	B	C
1	Measure Names	Month	DollarsTotal
2	Product Sales	Jan	9664
3	Product Sales	Feb	10446
4	Product Sales	Mar	9473
5	Product Sales	Apr	4119
6	Product Sales	May	150514
7	Product Sales	Jun	136145
8	Product Sales	Jul	154892
9	Product Sales	Aug	173479
10	Product Sales	Sep	40397
11	Product Sales	Oct	4040
12	Product Sales	Nov	3030

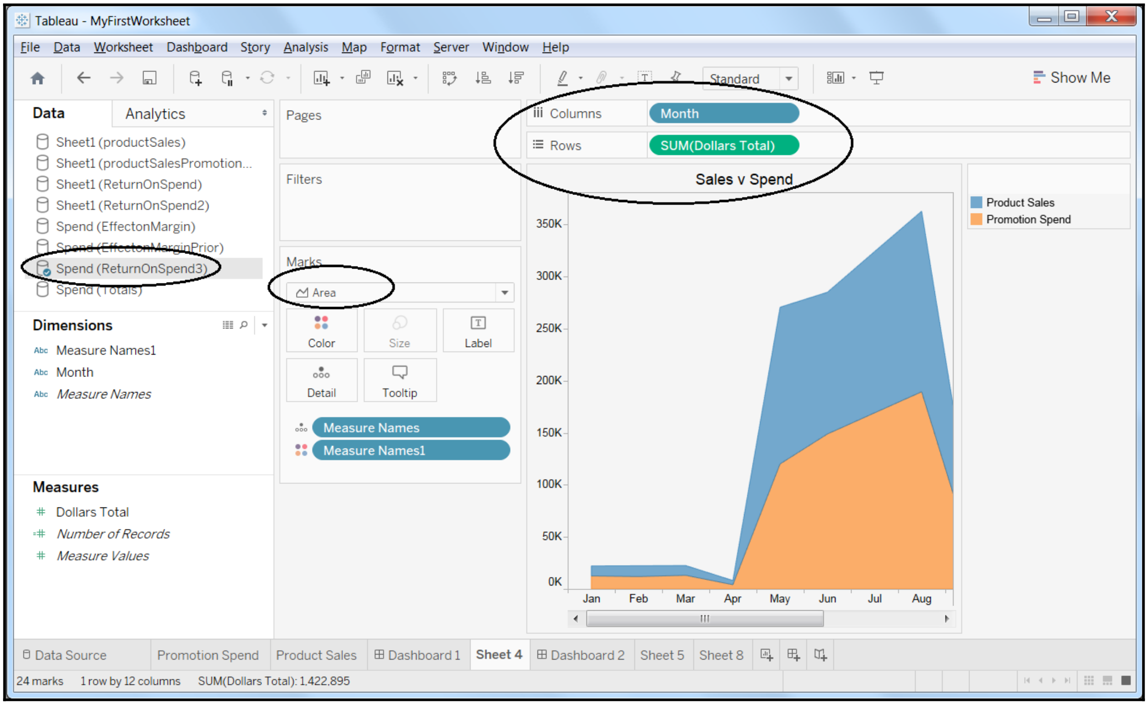


Tableau - MyFirstWorksheet

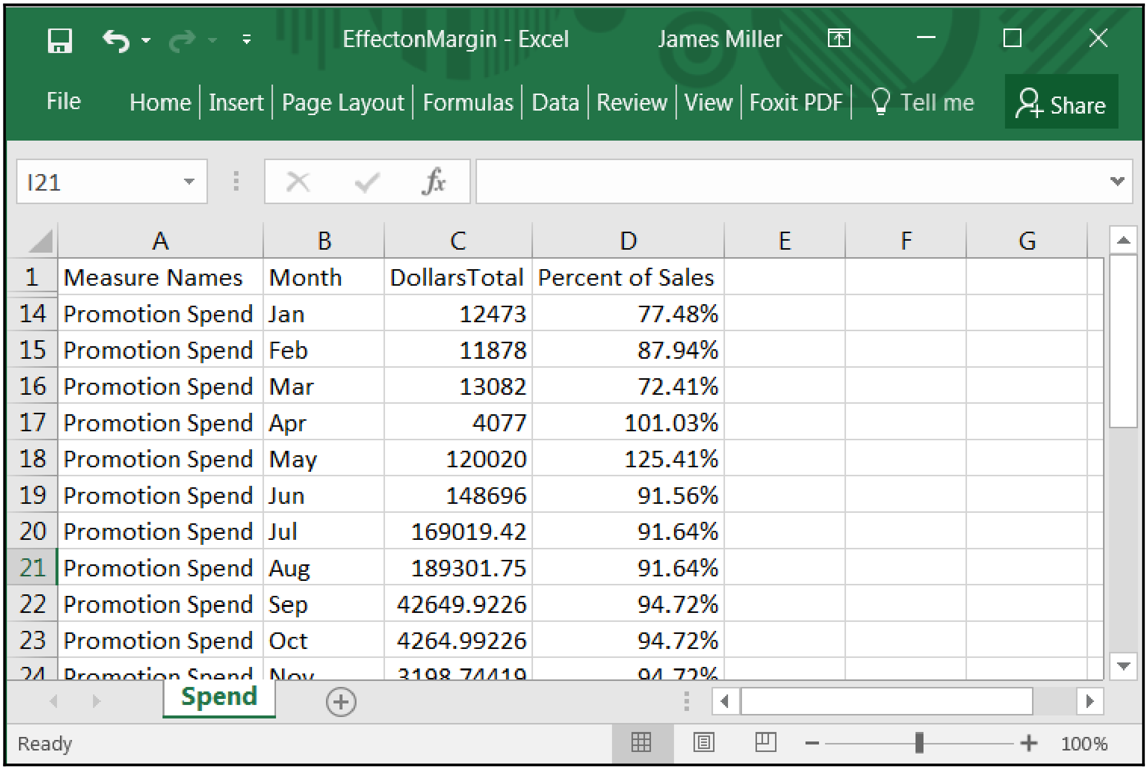
Columns: Month

Rows: SUM(Dollars Total)

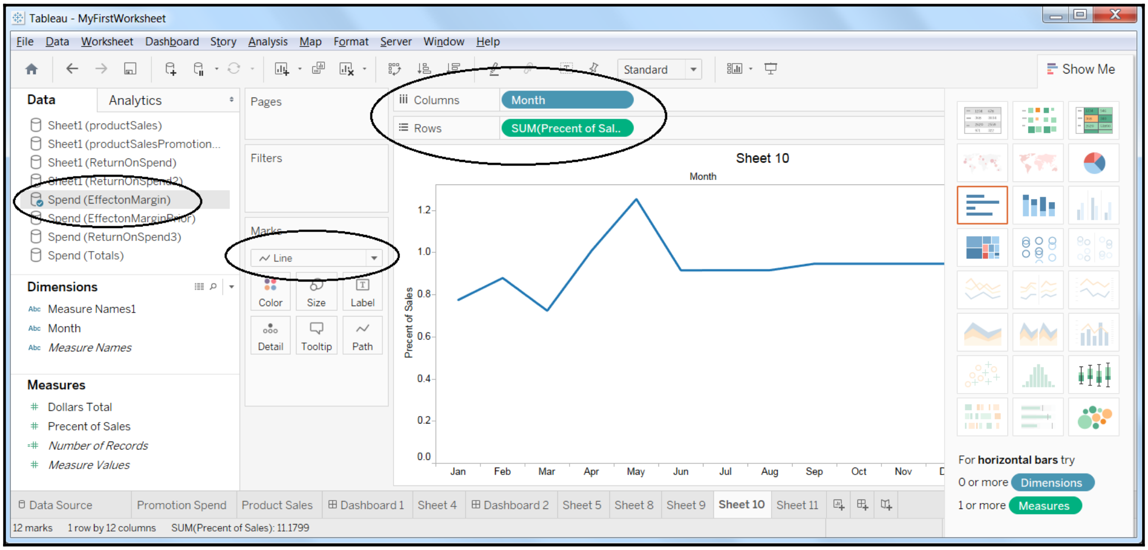
Marks: Area

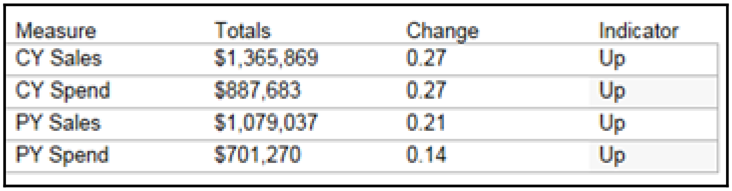
Sales v Spend

24 marks 1 row by 12 columns SUM(Dollars Total): 1,422,895

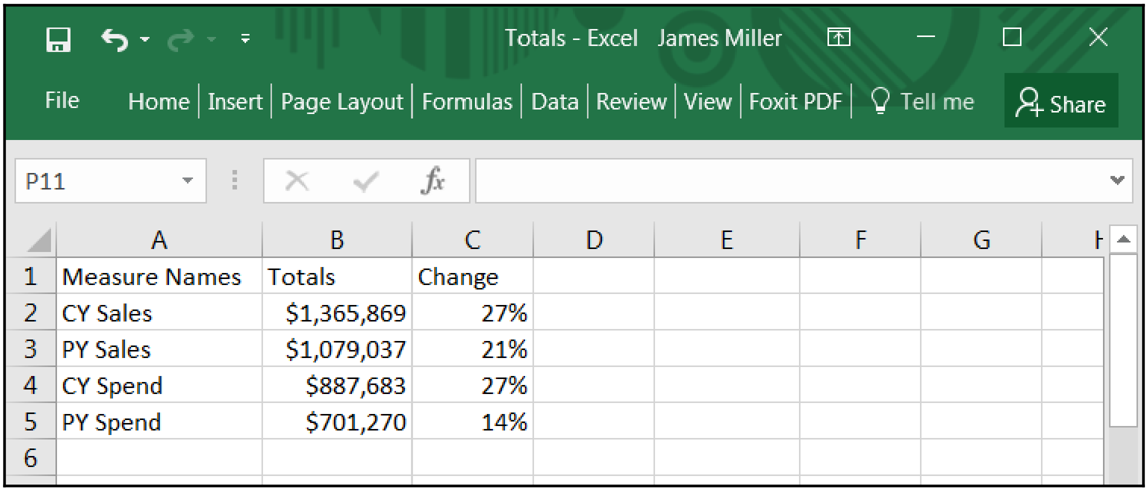


	A	B	C	D	E	F	G
1	Measure Names	Month	DollarsTotal	Percent of Sales			
14	Promotion Spend	Jan	12473	77.48%			
15	Promotion Spend	Feb	11878	87.94%			
16	Promotion Spend	Mar	13082	72.41%			
17	Promotion Spend	Apr	4077	101.03%			
18	Promotion Spend	May	120020	125.41%			
19	Promotion Spend	Jun	148696	91.56%			
20	Promotion Spend	Jul	169019.42	91.64%			
21	Promotion Spend	Aug	189301.75	91.64%			
22	Promotion Spend	Sep	42649.9226	94.72%			
23	Promotion Spend	Oct	4264.99226	94.72%			
24	Promotion Spend	Nov	3198.74419	94.72%			

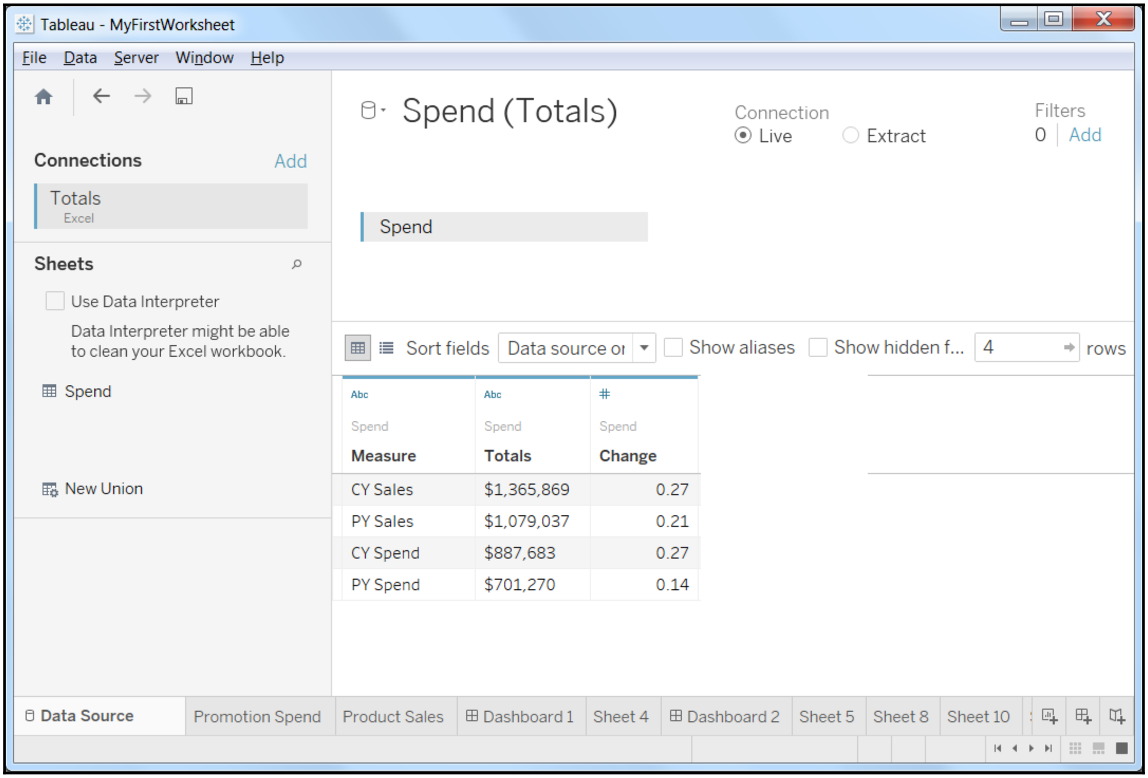




Measure	Totals	Change	Indicator
CY Sales	$1,365,869	0.27	Up
CY Spend	$887,683	0.27	Up
PY Sales	$1,079,037	0.21	Up
PY Spend	$701,270	0.14	Up



	A	B	C	D	E	F	G
1	Measure Names	Totals	Change				
2	CY Sales	$1,365,869	27%				
3	PY Sales	$1,079,037	21%				
4	CY Spend	$887,683	27%				
5	PY Spend	$701,270	14%				
6							



Spend (Totals)

Connection ● Live ○ Extract

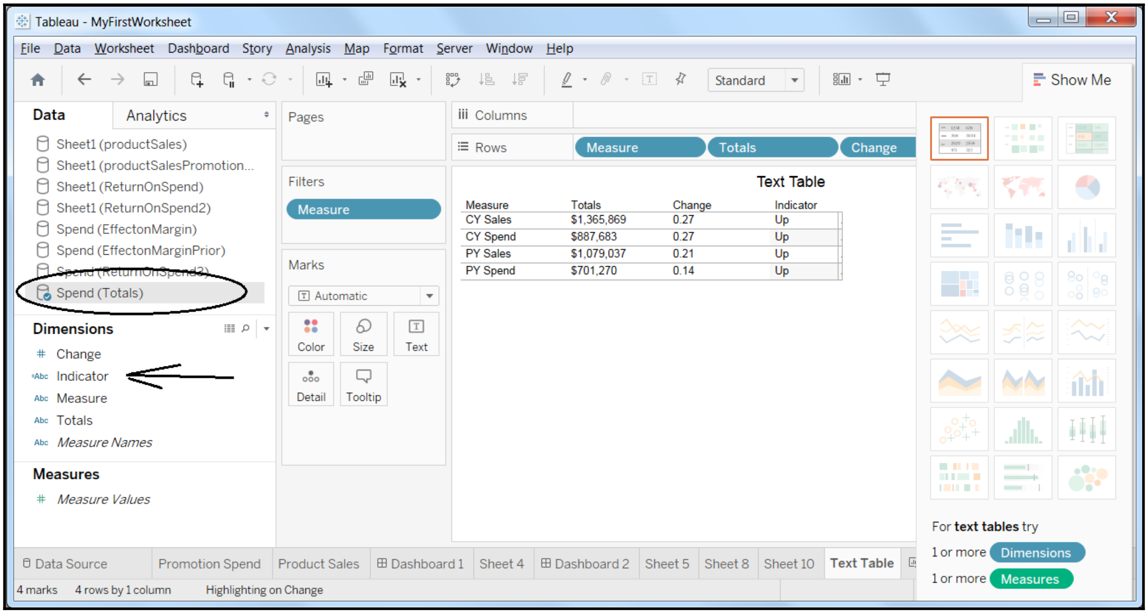
Filters 0 Add

□ Use Data Interpreter

Data Interpreter might be able to clean your Excel workbook.

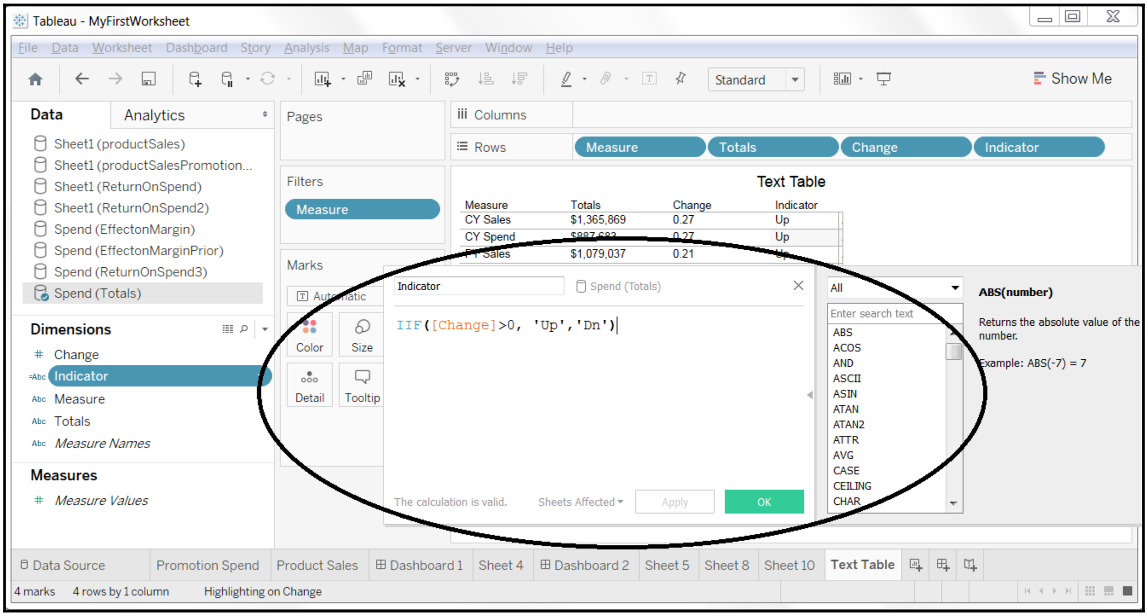
□ Show aliases □ Show hidden f... 4 rows

Abc Spend **Measure**	Abc Spend **Totals**	# Spend **Change**
CY Sales	$1,365,869	0.27
PY Sales	$1,079,037	0.21
CY Spend	$887,683	0.27
PY Spend	$701,270	0.14



Text Table

Measure	Totals	Change	Indicator
CY Sales	$1,365,869	0.27	Up
CY Spend	$887,683	0.27	Up
PY Sales	$1,079,037	0.21	Up
PY Spend	$701,270	0.14	Up



Text Table

Measure	Totals	Change	Indicator
CY Sales	$1,365,869	0.27	Up
CY Spend	$887,683	0.27	Up
PY Sales	$1,079,037	0.21	Up

Indicator   Spend (Totals)

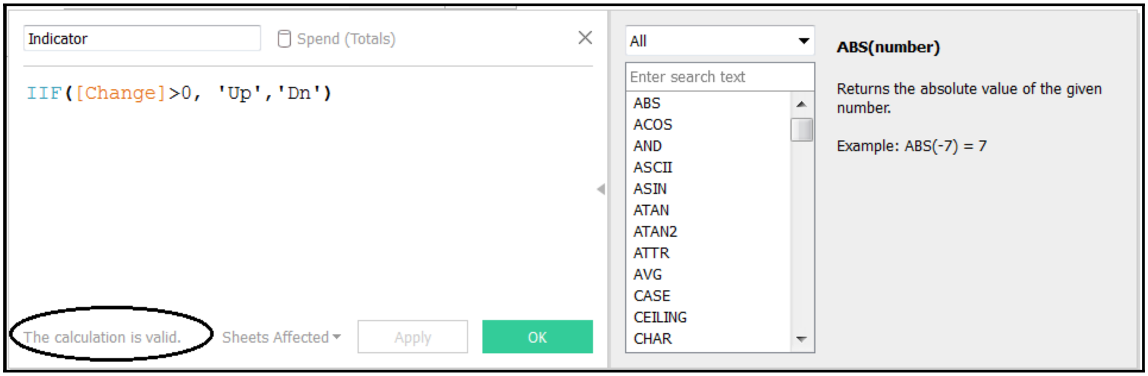
```
IIF([Change]>0, 'Up','Dn')
```

The calculation is valid.

ABS(number)

Returns the absolute value of the number.

Example: ABS(-7) = 7



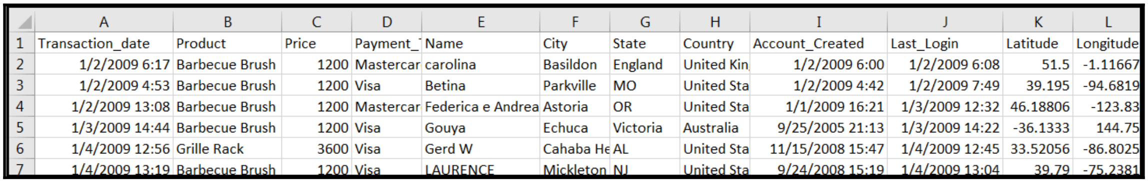
```
IIF([Change]>0, 'Up','Dn')
```

The calculation is valid.

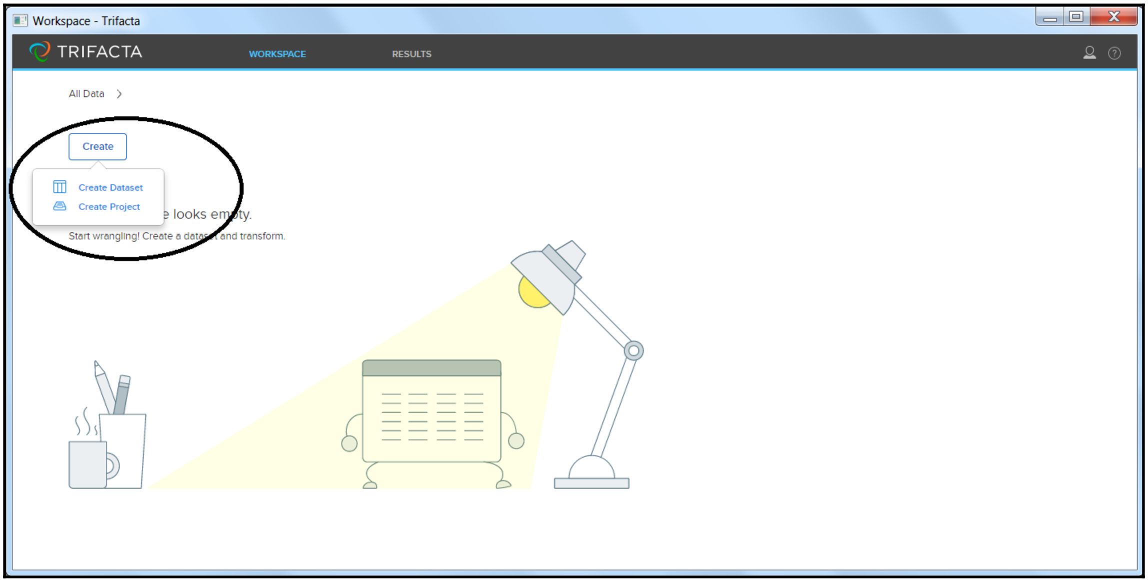
ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

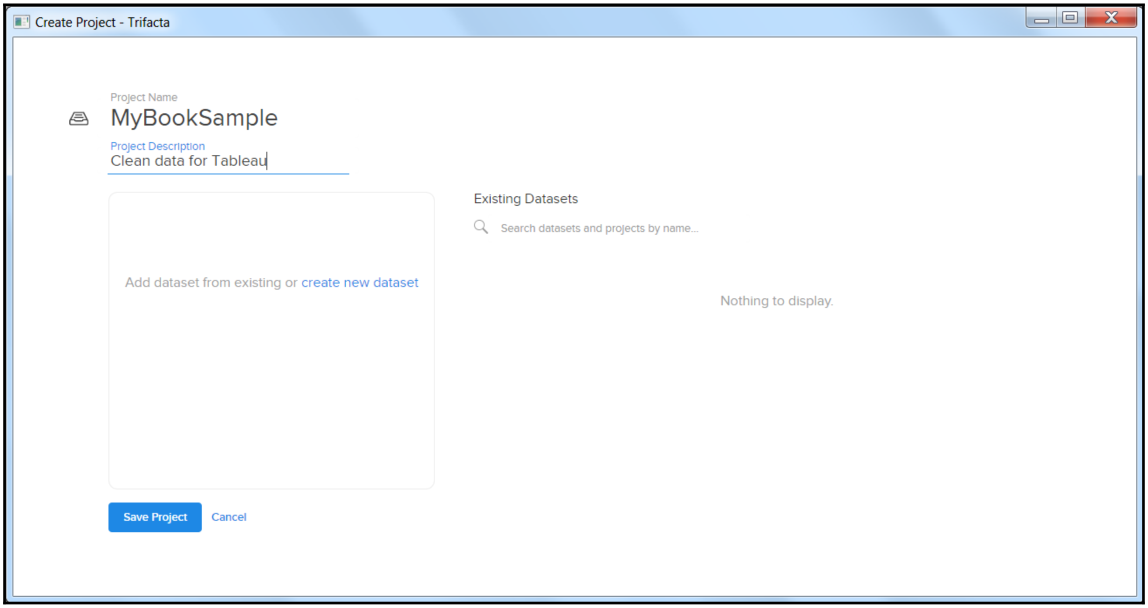


	A	B	C	D	E	F	G	H	I	J	K	L
1	Transaction_date	Product	Price	Payment_T	Name	City	State	Country	Account_Created	Last_Login	Latitude	Longitude
2	1/2/2009 6:17	Barbecue Brush	1200	Mastercar	carolina	Basildon	England	United Kin	1/2/2009 6:00	1/2/2009 6:08	51.5	-1.11667
3	1/2/2009 4:53	Barbecue Brush	1200	Visa	Betina	Parkville	MO	United Sta	1/2/2009 4:42	1/2/2009 7:49	39.195	-94.6819
4	1/2/2009 13:08	Barbecue Brush	1200	Mastercar	Federica e Andrea	Astoria	OR	United Sta	1/1/2009 16:21	1/3/2009 12:32	46.18806	-123.83
5	1/3/2009 14:44	Barbecue Brush	1200	Visa	Gouya	Echuca	Victoria	Australia	9/25/2005 21:13	1/3/2009 14:22	-36.1333	144.75
6	1/4/2009 12:56	Grille Rack	3600	Visa	Gerd W	Cahaba He	AL	United Sta	11/15/2008 15:47	1/4/2009 12:45	33.52056	-86.8025
7	1/4/2009 13:19	Barbecue Brush	1200	Visa	LAURENCE	Mickleton	NJ	United Sta	9/24/2008 15:19	1/4/2009 13:04	39.79	-75.2381



e looks empty.

Start wrangling! Create a dataset and transform.



Create Project - Trifacta

Project Name
MyBookSample

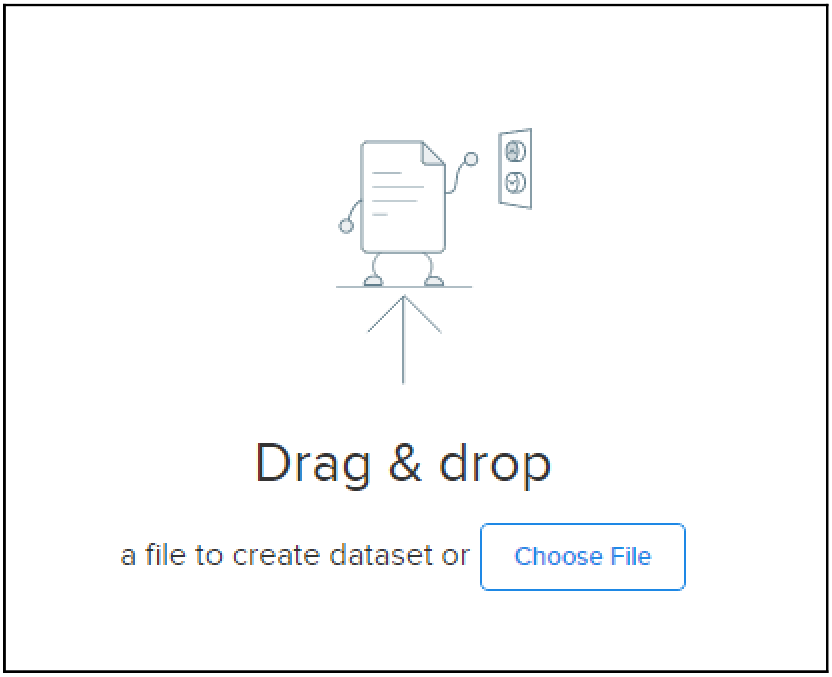
Project Description
Clean data for Tableau

Add dataset from existing or create new dataset

Existing Datasets

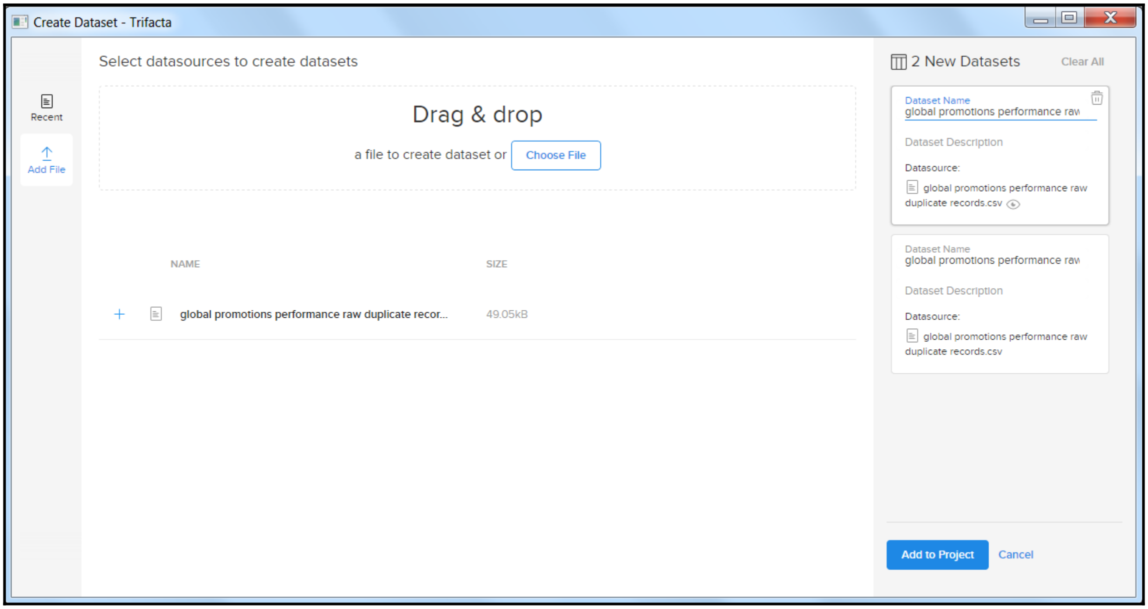
Search datasets and projects by name...

Nothing to display.



Drag & drop

a file to create dataset or



Select datasources to create datasets

Drag & drop

a file to create dataset or Choose File

NAME	SIZE
global promotions performance raw duplicate recor...	49.05kB

2 New Datasets

Dataset Name
global promotions performance raw

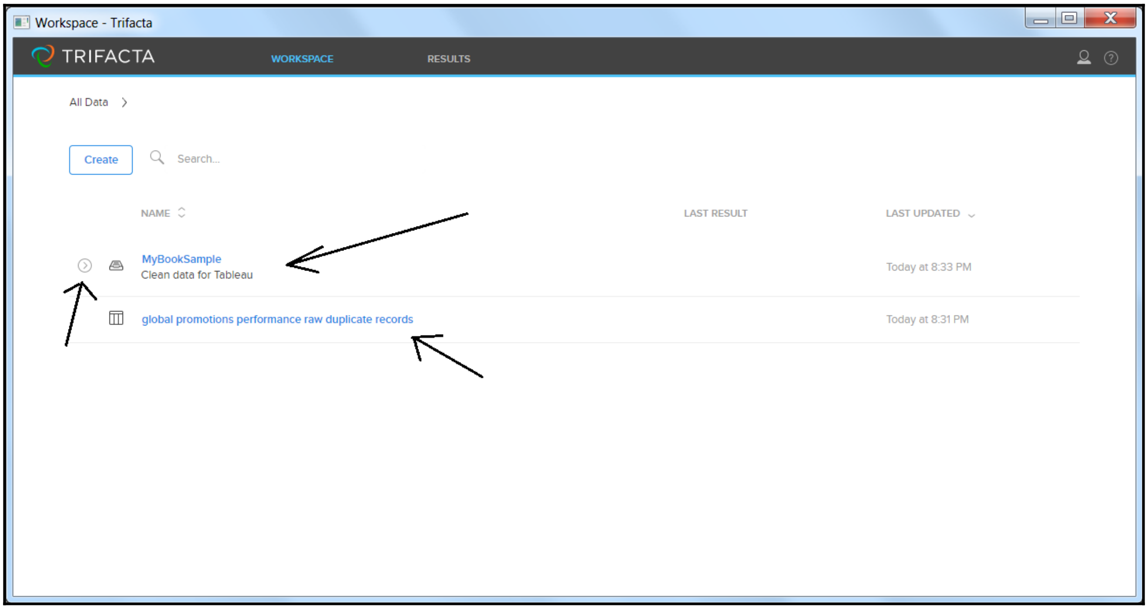
Dataset Description

Datasource:
global promotions performance raw duplicate records.csv

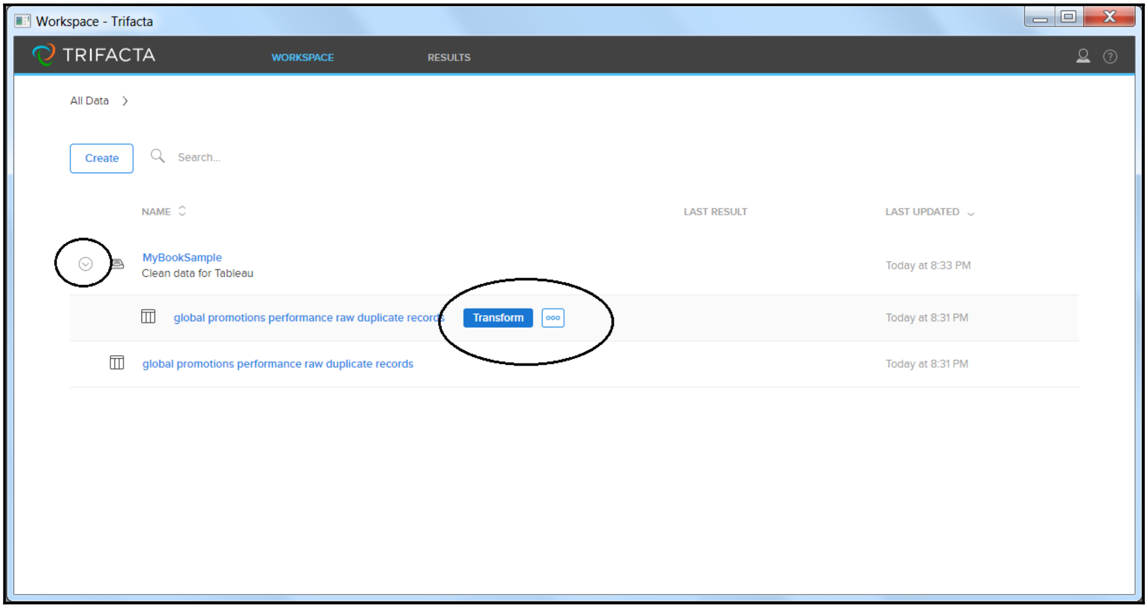
Dataset Name
global promotions performance raw

Dataset Description

Datasource:
global promotions performance raw duplicate records.csv



NAME	LAST RESULT	LAST UPDATED
MyBookSample Clean data for Tableau		Today at 8:33 PM
global promotions performance raw duplicate records		Today at 8:31 PM



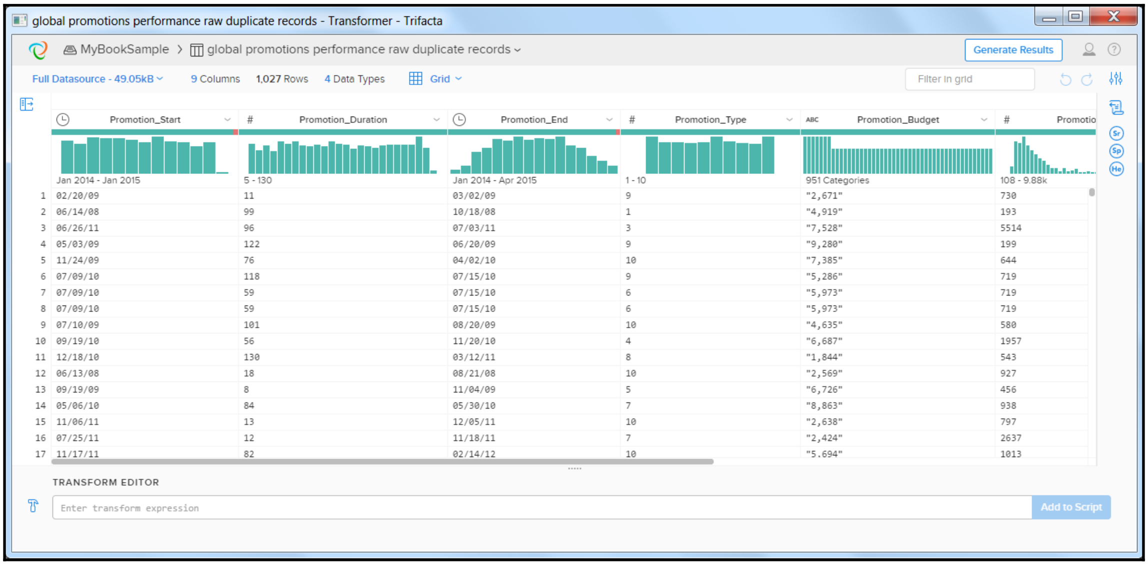
Workspace - Trifacta

TRIFACTA WORKSPACE RESULTS

All Data >

Create Search...

NAME	LAST RESULT	LAST UPDATED
MyBookSample Clean data for Tableau		Today at 8:33 PM
global promotions performance raw duplicate records Transform ooo		Today at 8:31 PM
global promotions performance raw duplicate records		Today at 8:31 PM



global promotions performance raw duplicate records - Transformer - Trifacta

MyBookSample > global promotions performance raw duplicate records

Generate Results

Full Datasource - 49.05kB 9 Columns 1,027 Rows 4 Data Types Grid

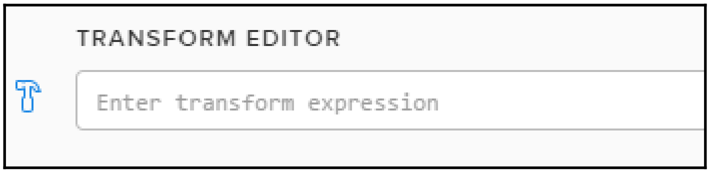
Filter in grid

	Promotion_Start	Promotion_Duration	Promotion_End	Promotion_Type	Promotion_Budget	Promotio
	Jan 2014 - Jan 2015	5 - 130	Jan 2014 - Apr 2015	1 - 10	951 Categories	108 - 9.88k
1	02/20/09	11	03/02/09	9	"2,671"	730
2	06/14/08	99	10/18/08	1	"4,919"	193
3	06/26/11	96	07/03/11	3	"7,528"	5514
4	05/03/09	122	06/20/09	9	"9,280"	199
5	11/24/09	76	04/02/10	10	"7,385"	644
6	07/09/10	118	07/15/10	9	"5,286"	719
7	07/09/10	59	07/15/10	6	"5,973"	719
8	07/09/10	59	07/15/10	6	"5,973"	719
9	07/10/09	101	08/20/09	10	"4,635"	580
10	09/19/10	56	11/20/10	4	"6,687"	1957
11	12/18/10	130	03/12/11	8	"1,844"	543
12	06/13/08	18	08/21/08	10	"2,569"	927
13	09/19/09	8	11/04/09	5	"6,726"	456
14	05/06/10	84	05/30/10	7	"8,863"	938
15	11/06/11	13	12/05/11	10	"2,638"	797
16	07/25/11	12	11/18/11	7	"2,424"	2637
17	11/17/11	82	02/14/12	10	"5.694"	1013

TRANSFORM EDITOR

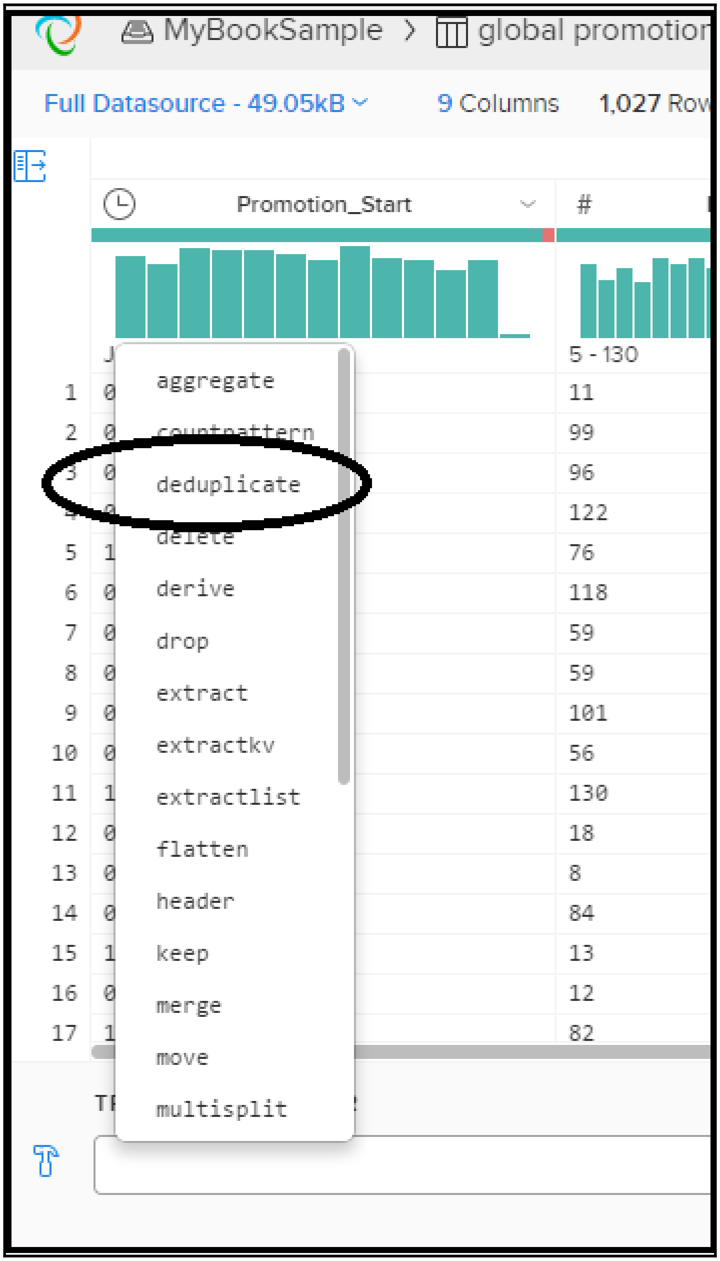
Enter transform expression

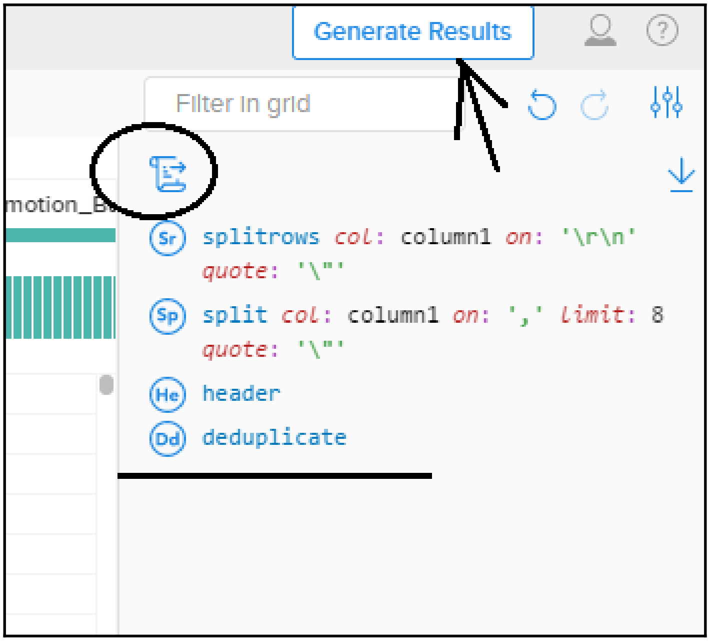
Add to Script



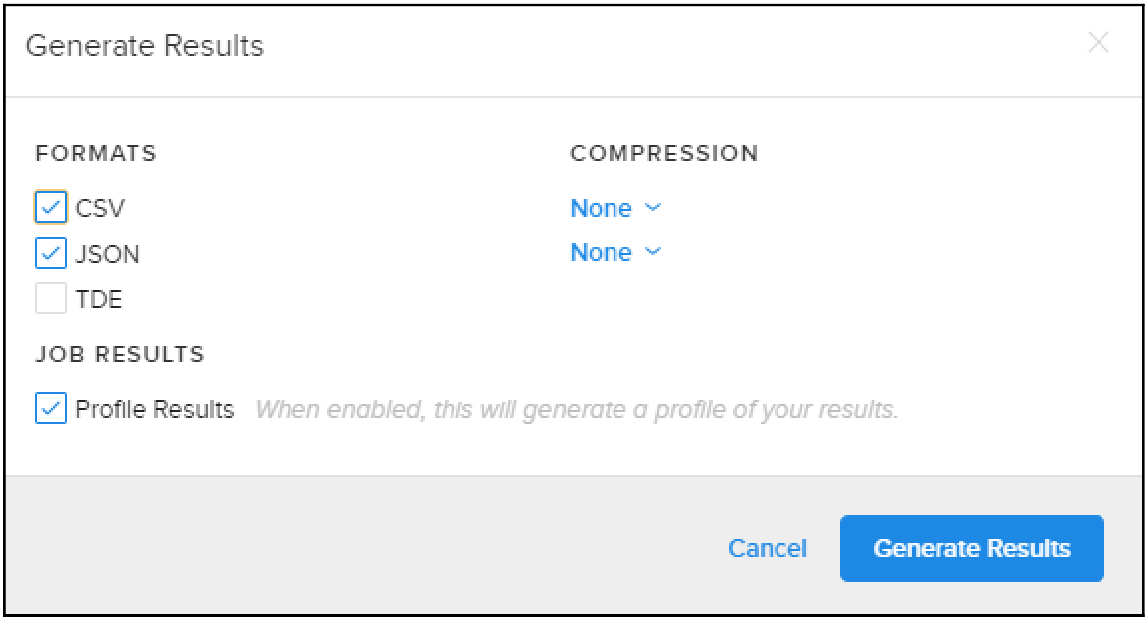
TRANSFORM EDITOR

Enter transform expression


MyBookSample
global promotion
Full Datasource - 49.05kB
9 Columns
1,027 Row
Promotion_Start
#
5 - 130
aggregate
countpattern
deduplicate
delete
derive
drop
extract
extractkv
extractlist
flatten
header
keep
merge
move
multisplit
1 0 11
2 0 99
3 0 96
4 0 122
5 1 76
6 0 118
7 0 59
8 0 59
9 0 101
10 0 56
11 1 130
12 0 18
13 0 8
14 0 84
15 1 13
16 0 12
17 1 82



```
Sr splitrows col: column1 on: '\r\n'
   quote: '\"'
Sp split col: column1 on: ',' limit: 8
   quote: '\"'
He header
Dd deduplicate
```



Generate Results

FORMATS

☑ CSV

☑ JSON

□ TDE

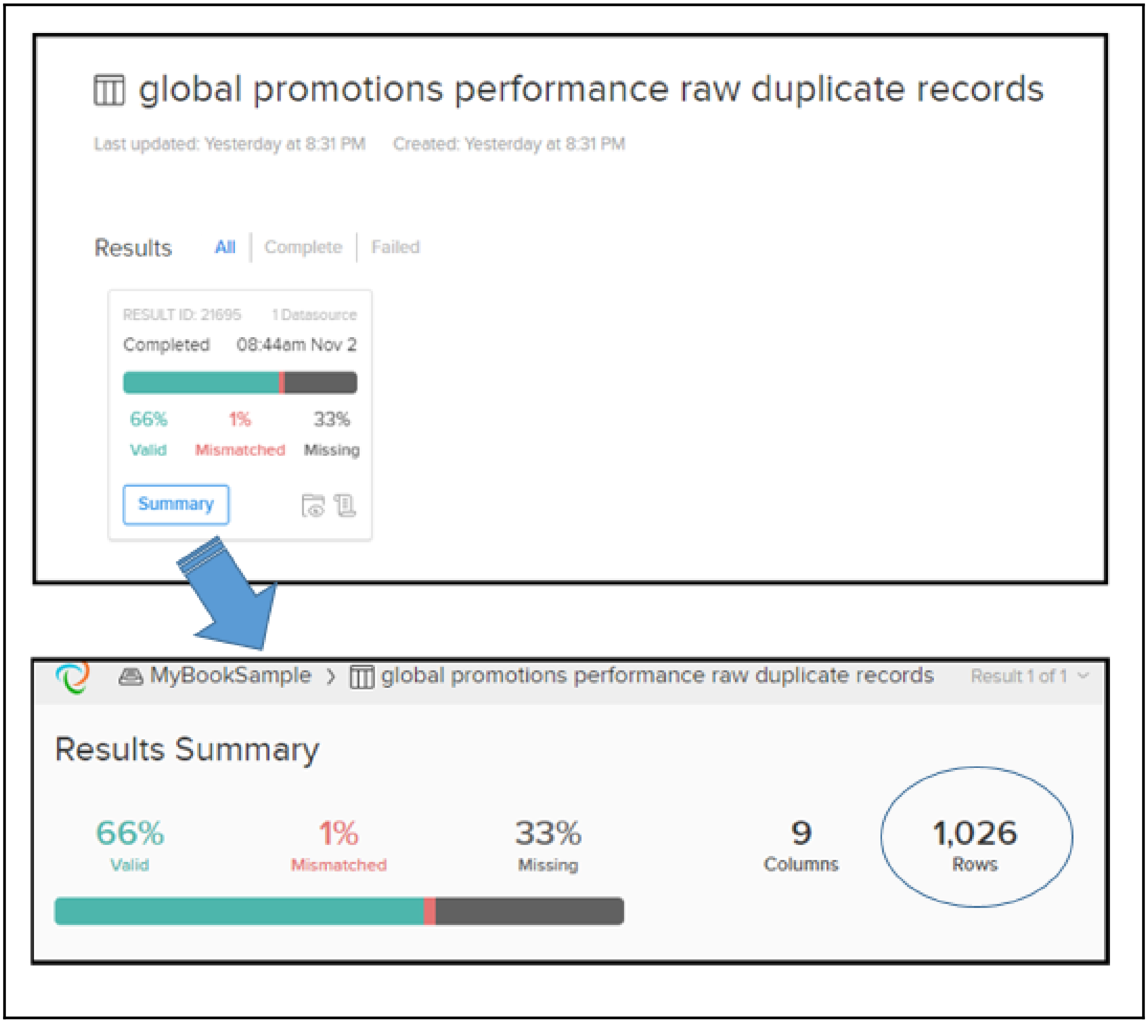
COMPRESSION

None

None

JOB RESULTS

☑ Profile Results *When enabled, this will generate a profile of your results.*



global promotions performance raw duplicate records

Last updated: Yesterday at 8:31 PM Created: Yesterday at 8:31 PM

Results All | Complete | Failed

RESULT ID: 21695 1 Datasource

Completed 08:44am Nov 2

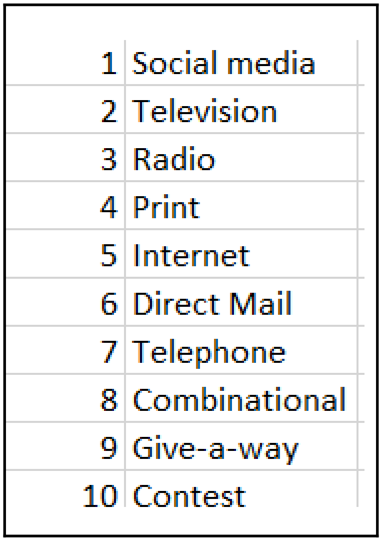
66% Valid 1% Mismatched 33% Missing

Summary

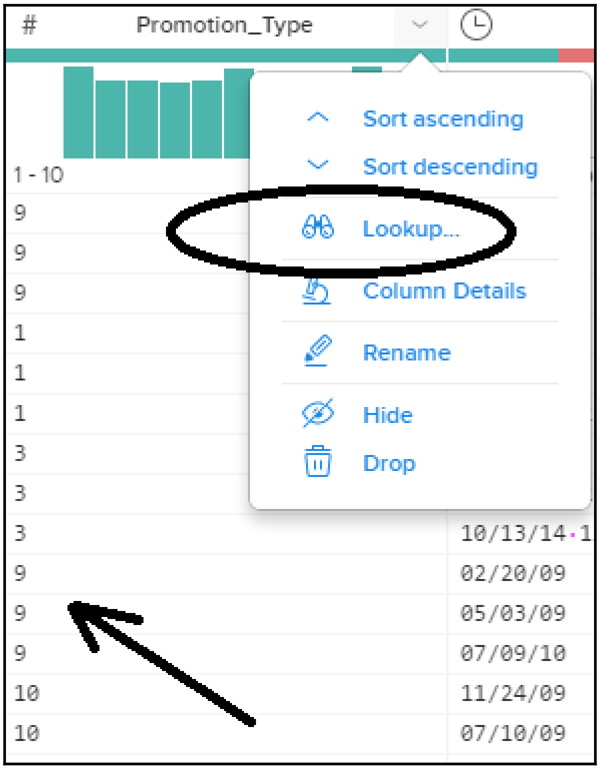
MyBookSample › global promotions performance raw duplicate records Result 1 of 1

Results Summary

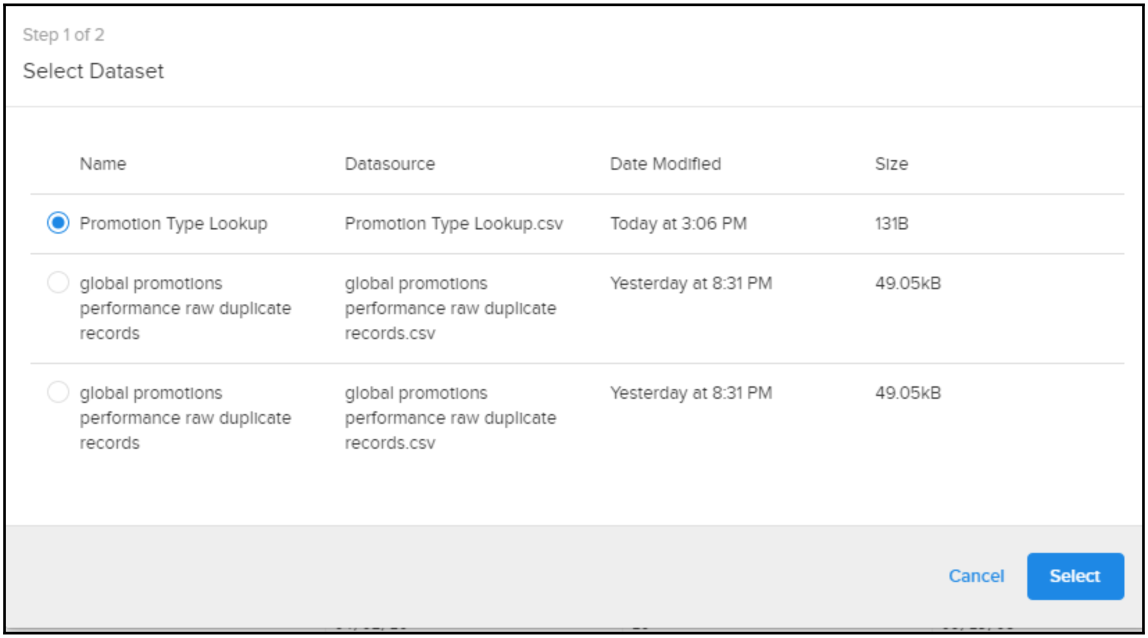
66%	1%	33%	9	1,026
Valid	Mismatched	Missing	Columns	Rows



1	Social media
2	Television
3	Radio
4	Print
5	Internet
6	Direct Mail
7	Telephone
8	Combinational
9	Give-a-way
10	Contest



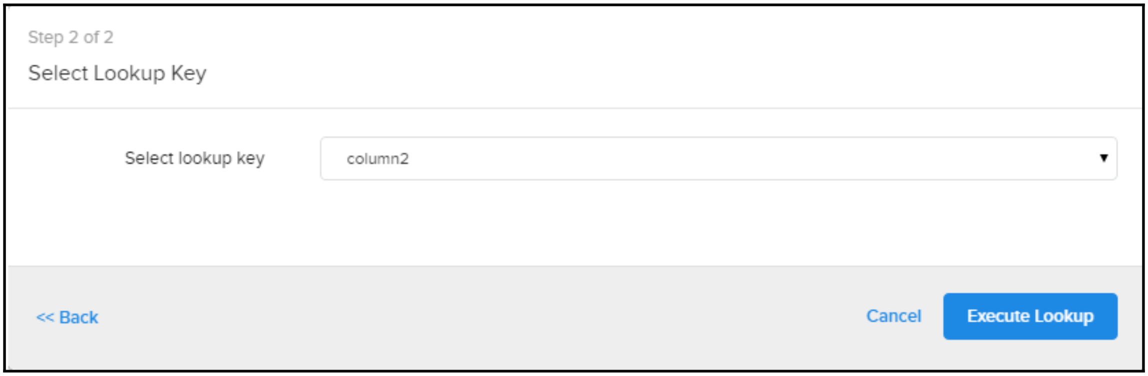
# Promotion_Type	
9	
9	
9	
1	
1	
1	
3	
3	
3	10/13/14·1
9	02/20/09
9	05/03/09
9	07/09/10
10	11/24/09
10	07/10/09



Step 1 of 2

Select Dataset

	Name	Datasource	Date Modified	Size
●	Promotion Type Lookup	Promotion Type Lookup.csv	Today at 3:06 PM	131B
○	global promotions performance raw duplicate records	global promotions performance raw duplicate records.csv	Yesterday at 8:31 PM	49.05kB
○	global promotions performance raw duplicate records	global promotions performance raw duplicate records.csv	Yesterday at 8:31 PM	49.05kB



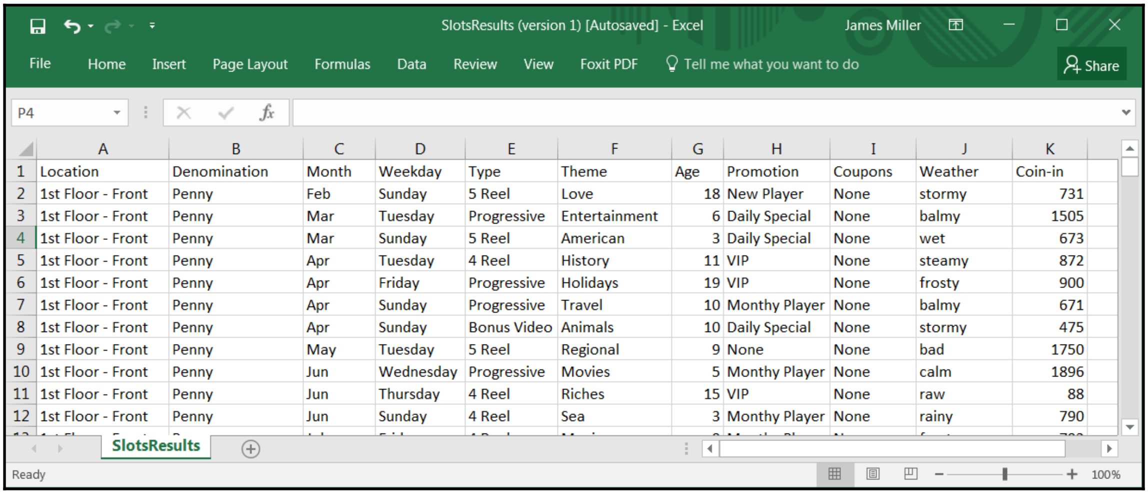
Step 2 of 2

Select Lookup Key

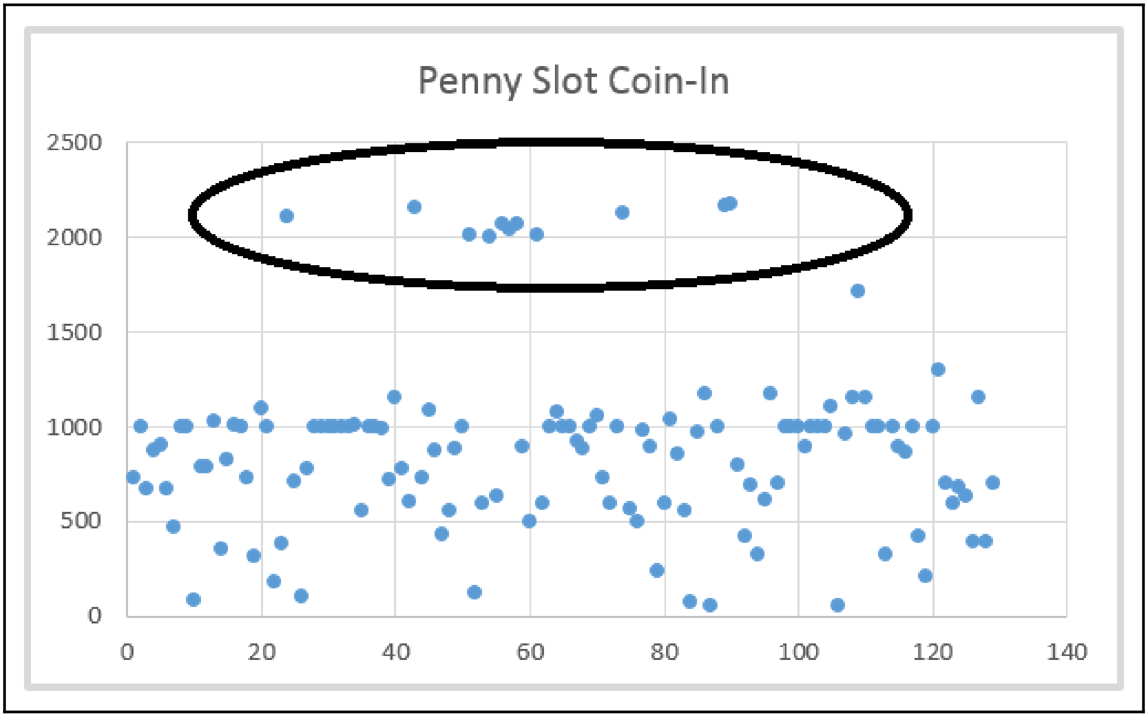
Select lookup key: column2 ▾

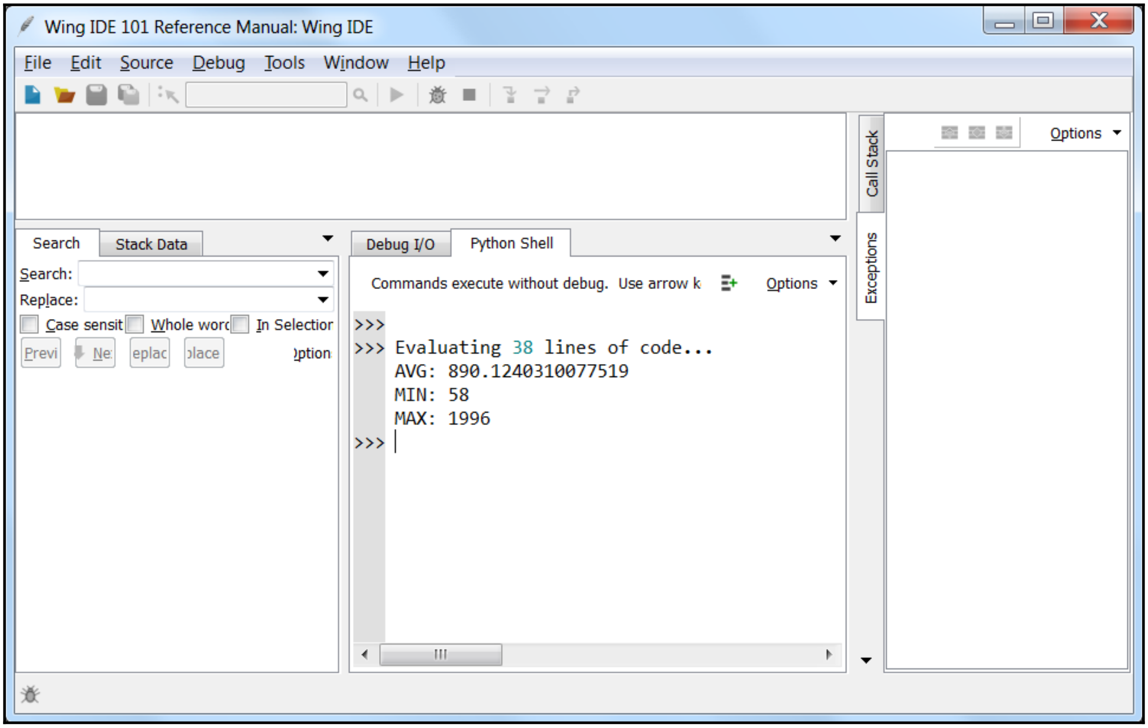
Chapter 7: Dealing with Outliers Using Python

WINGWARE
PYTHON IDE
101 v. 5.1
THE INTELLIGENT DEVELOPMENT ENVIRONMENT for PYTHON PROGRAMMERS
Copyright 1999-2015, Archaeopteryx Software, Inc d.b.a. Wingware



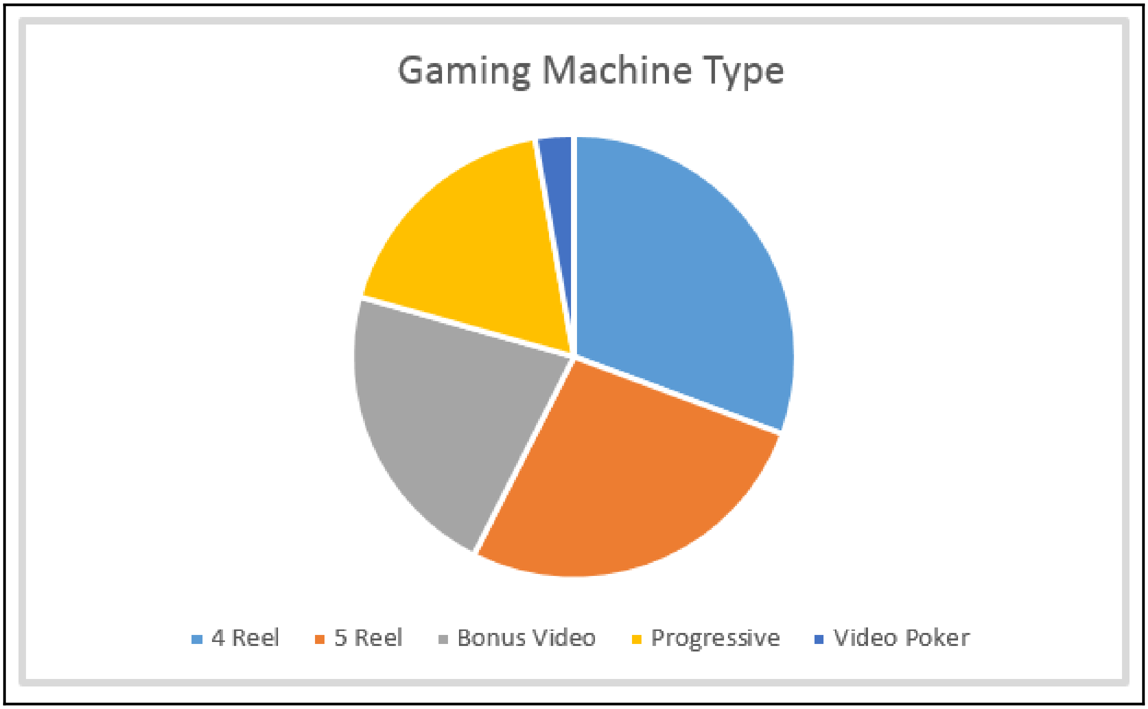
	A	B	C	D	E	F	G	H	I	J	K
1	Location	Denomination	Month	Weekday	Type	Theme	Age	Promotion	Coupons	Weather	Coin-in
2	1st Floor - Front	Penny	Feb	Sunday	5 Reel	Love	18	New Player	None	stormy	731
3	1st Floor - Front	Penny	Mar	Tuesday	Progressive	Entertainment	6	Daily Special	None	balmy	1505
4	1st Floor - Front	Penny	Mar	Sunday	5 Reel	American	3	Daily Special	None	wet	673
5	1st Floor - Front	Penny	Apr	Tuesday	4 Reel	History	11	VIP	None	steamy	872
6	1st Floor - Front	Penny	Apr	Friday	Progressive	Holidays	19	VIP	None	frosty	900
7	1st Floor - Front	Penny	Apr	Sunday	Progressive	Travel	10	Monthy Player	None	balmy	671
8	1st Floor - Front	Penny	Apr	Sunday	Bonus Video	Animals	10	Daily Special	None	stormy	475
9	1st Floor - Front	Penny	May	Tuesday	5 Reel	Regional	9	None	None	bad	1750
10	1st Floor - Front	Penny	Jun	Wednesday	Progressive	Movies	5	Monthy Player	None	calm	1896
11	1st Floor - Front	Penny	Jun	Thursday	4 Reel	Riches	15	VIP	None	raw	88
12	1st Floor - Front	Penny	Jun	Sunday	4 Reel	Sea	3	Monthy Player	None	rainy	790

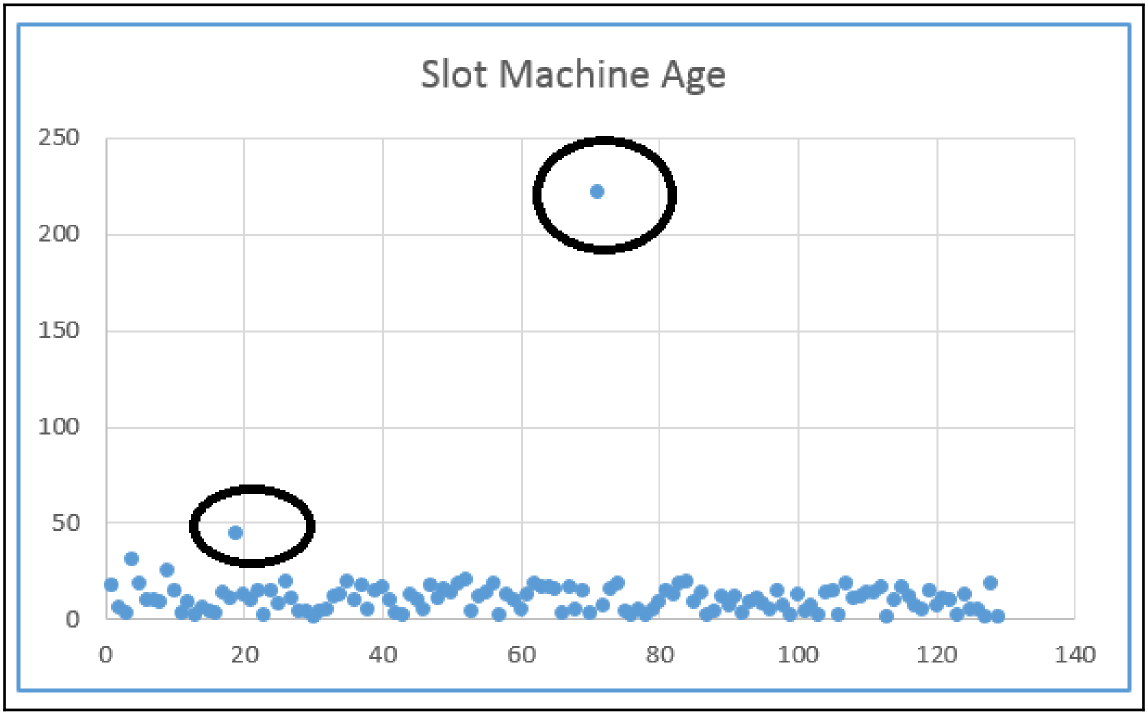

Penny Slot Coin-In
2500
2000
1500
1000
500
0
0
20
40
60
80
100
120
140

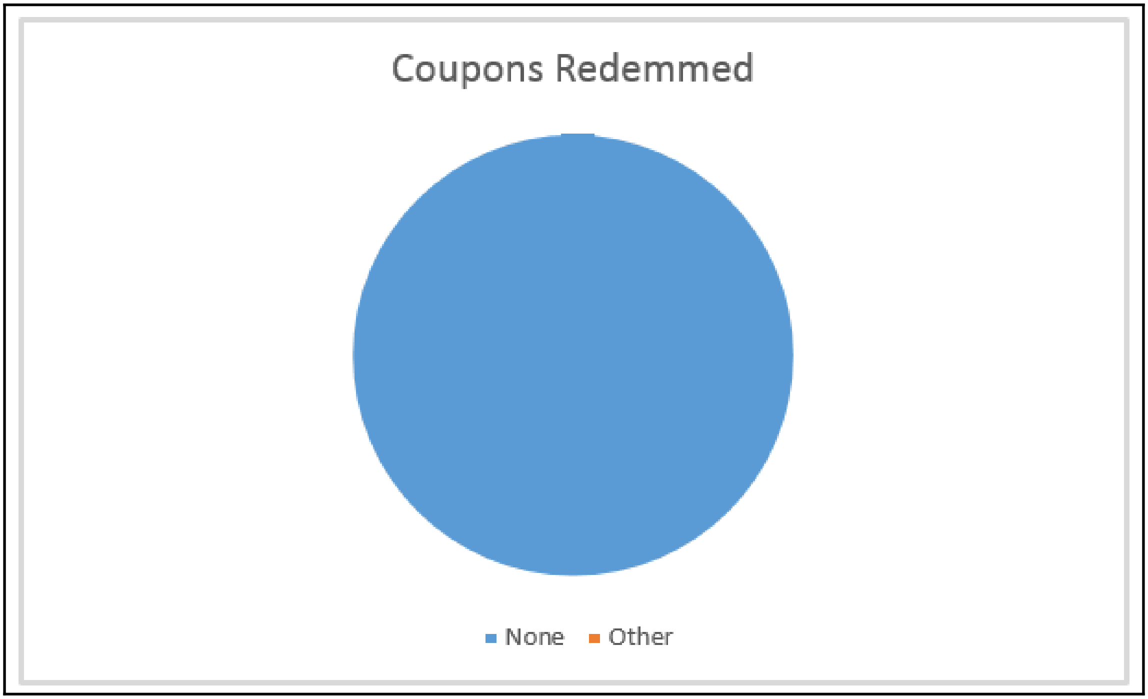


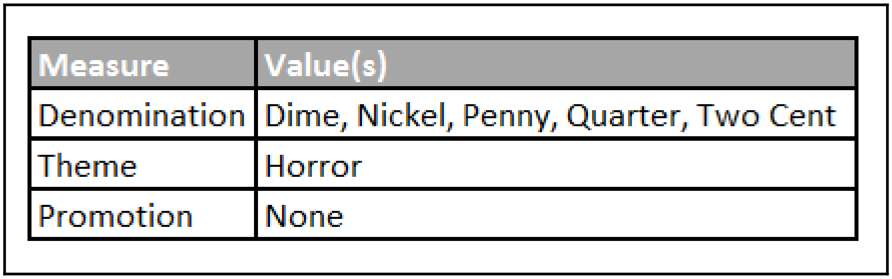
Commands execute without debug. Use arrow k

```
>>>
>>> Evaluating 38 lines of code...
    AVG: 890.1240310077519
    MIN: 58
    MAX: 1996
>>>
```


Gaming Machine Type
4 Reel
5 Reel
Bonus Video
Progressive
Video Poker

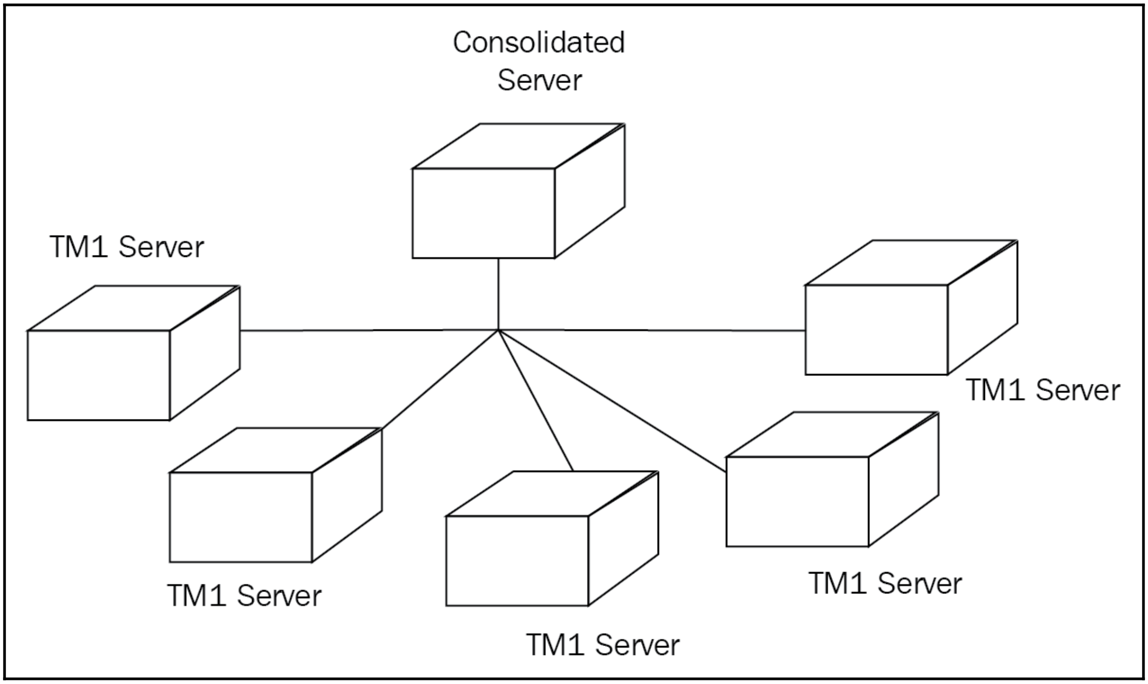

Slot Machine Age
250
200
150
100
50
0
0
20
40
60
80
100
120
140

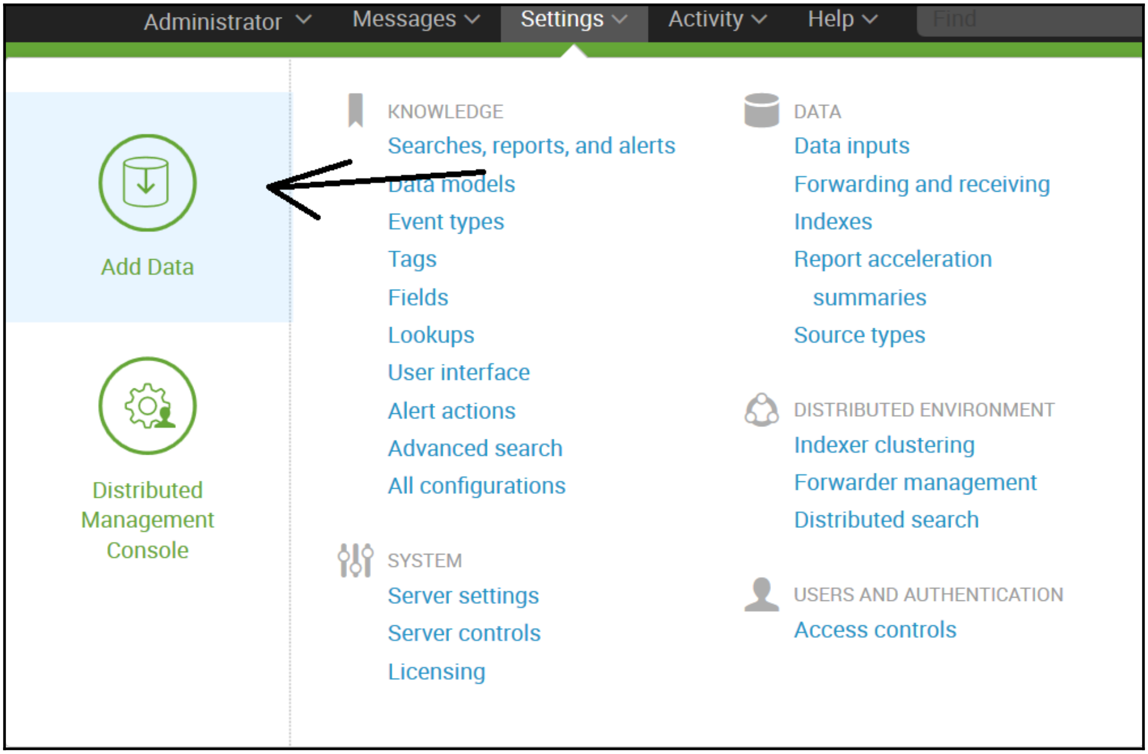


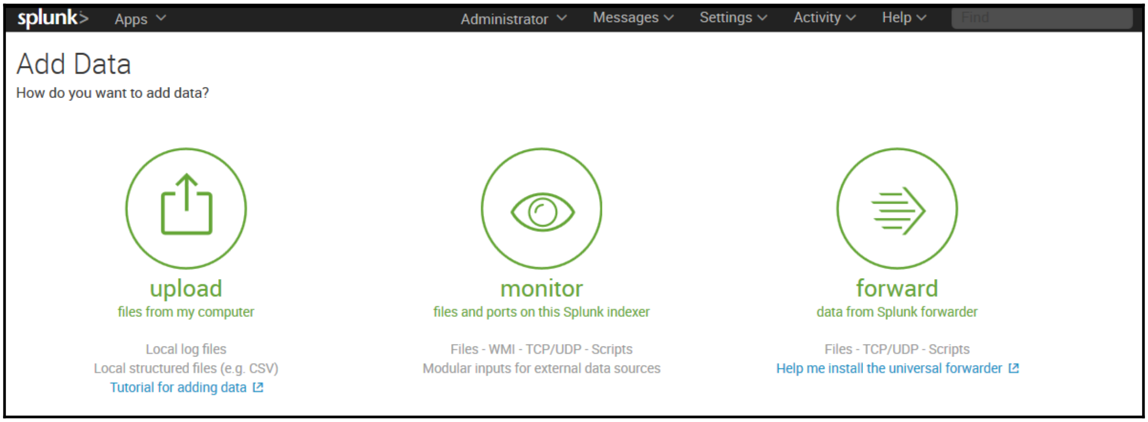


Measure	Value(s)
Denomination	Dime, Nickel, Penny, Quarter, Two Cent
Theme	Horror
Promotion	None

Chapter 8: Big Data Operational Intelligence with Splunk







Add Data

How do you want to add data?

upload
files from my computer

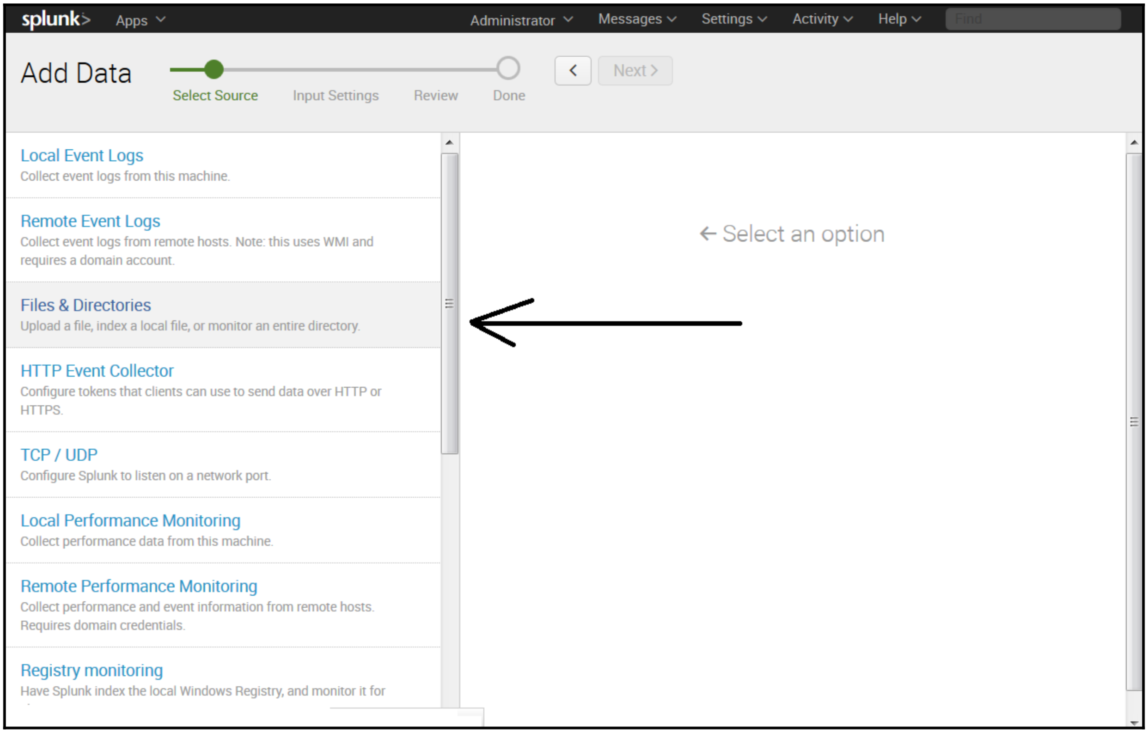
Local log files
Local structured files (e.g. CSV)
Tutorial for adding data

monitor
files and ports on this Splunk indexer

Files - WMI - TCP/UDP - Scripts
Modular inputs for external data sources

forward
data from Splunk forwarder

Files - TCP/UDP - Scripts
Help me install the universal forwarder



Local Event Logs
Collect event logs from this machine.

Remote Event Logs
Collect event logs from remote hosts. Note: this uses WMI and requires a domain account.

Files & Directories
Upload a file, index a local file, or monitor an entire directory.

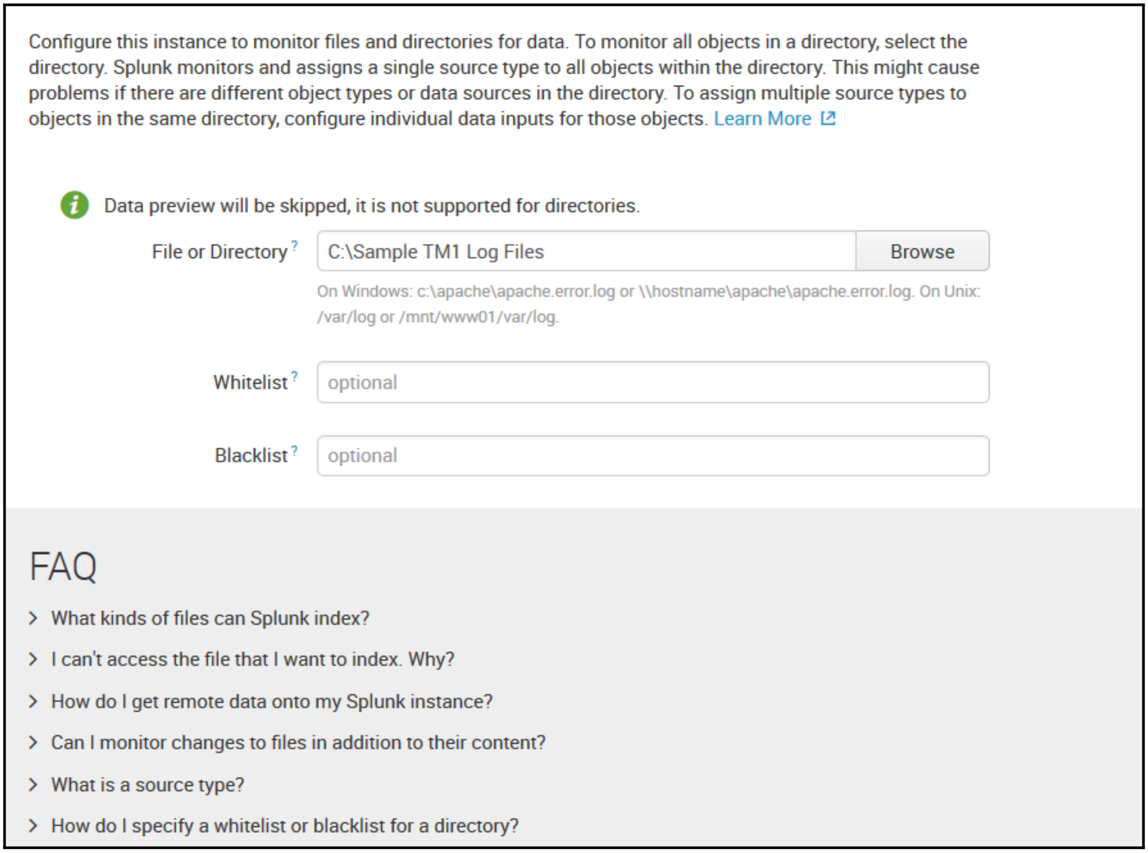
HTTP Event Collector
Configure tokens that clients can use to send data over HTTP or HTTPS.

TCP / UDP
Configure Splunk to listen on a network port.

Local Performance Monitoring
Collect performance data from this machine.

Remote Performance Monitoring
Collect performance and event information from remote hosts. Requires domain credentials.

Registry monitoring
Have Splunk index the local Windows Registry, and monitor it for



Configure this instance to monitor files and directories for data. To monitor all objects in a directory, select the directory. Splunk monitors and assigns a single source type to all objects within the directory. This might cause problems if there are different object types or data sources in the directory. To assign multiple source types to objects in the same directory, configure individual data inputs for those objects. Learn More

Data preview will be skipped, it is not supported for directories.

File or Directory ? C:\Sample TM1 Log Files

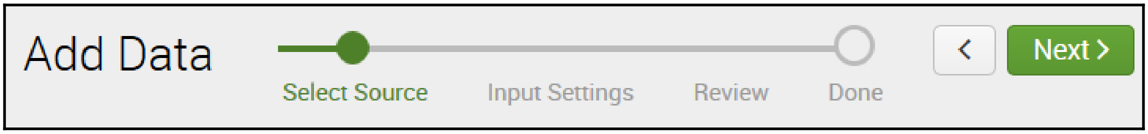
On Windows: c:\apache\apache.error.log or \\hostname\apache\apache.error.log. On Unix: /var/log or /mnt/www01/var/log.

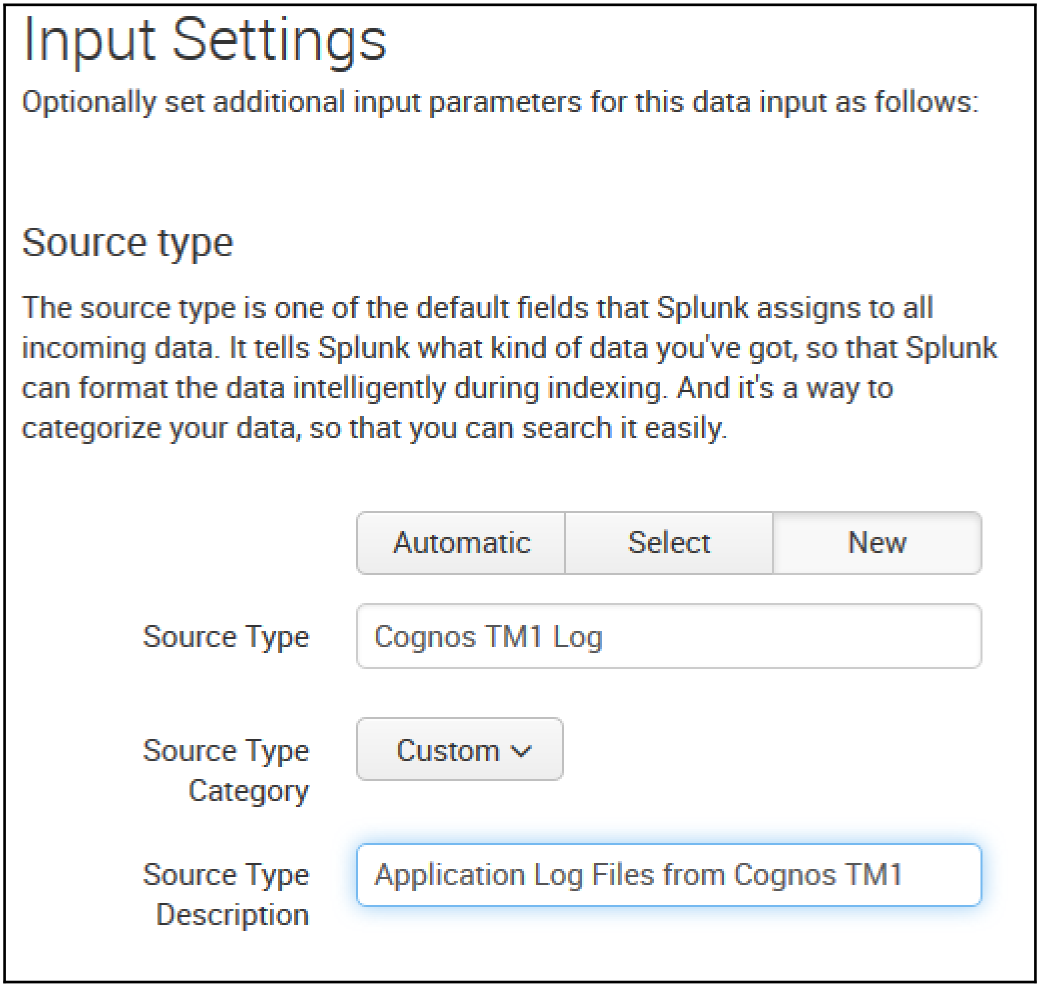
Whitelist ? optional

Blacklist ? optional

FAQ

- What kinds of files can Splunk index?
- I can't access the file that I want to index. Why?
- How do I get remote data onto my Splunk instance?
- Can I monitor changes to files in addition to their content?
- What is a source type?
- How do I specify a whitelist or blacklist for a directory?





Input Settings

Optionally set additional input parameters for this data input as follows:

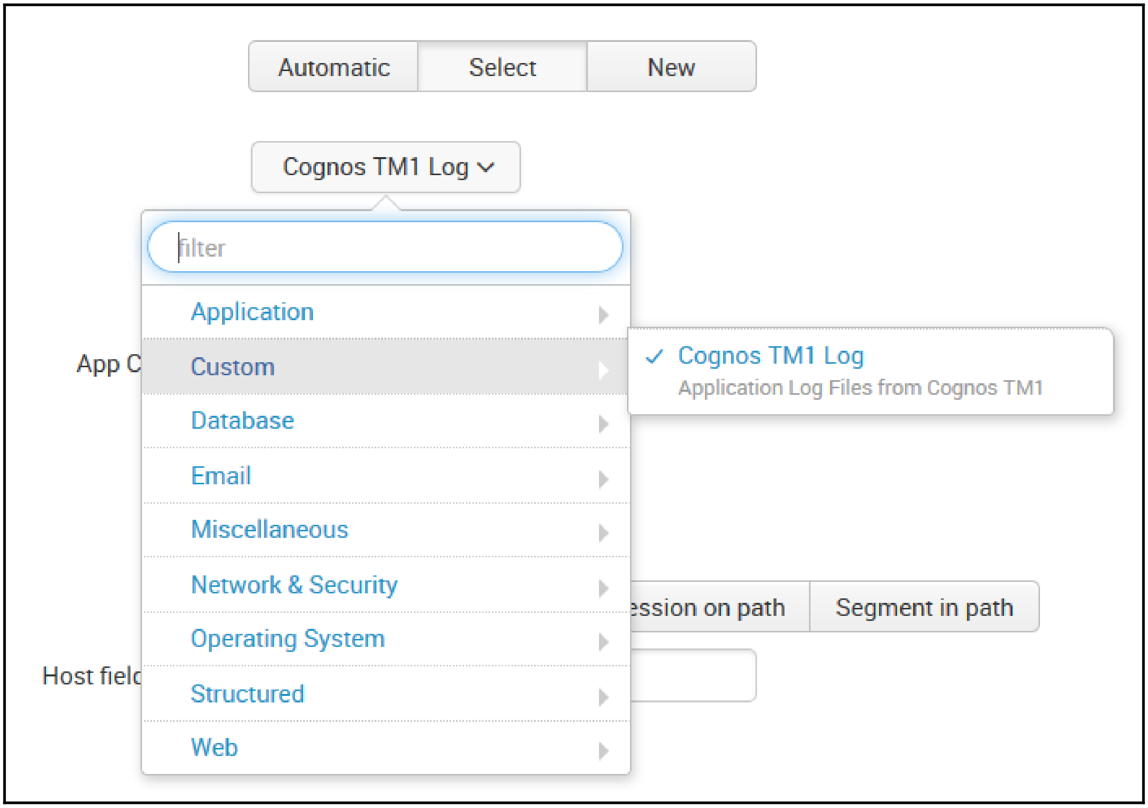
Source type

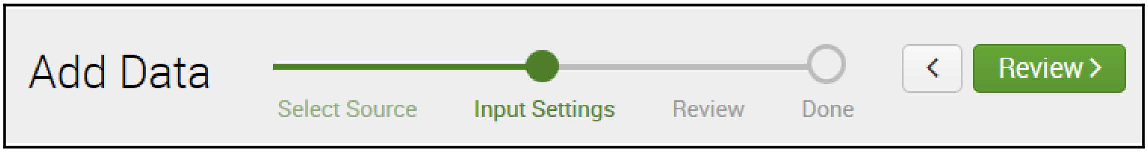
The source type is one of the default fields that Splunk assigns to all incoming data. It tells Splunk what kind of data you've got, so that Splunk can format the data intelligently during indexing. And it's a way to categorize your data, so that you can search it easily.

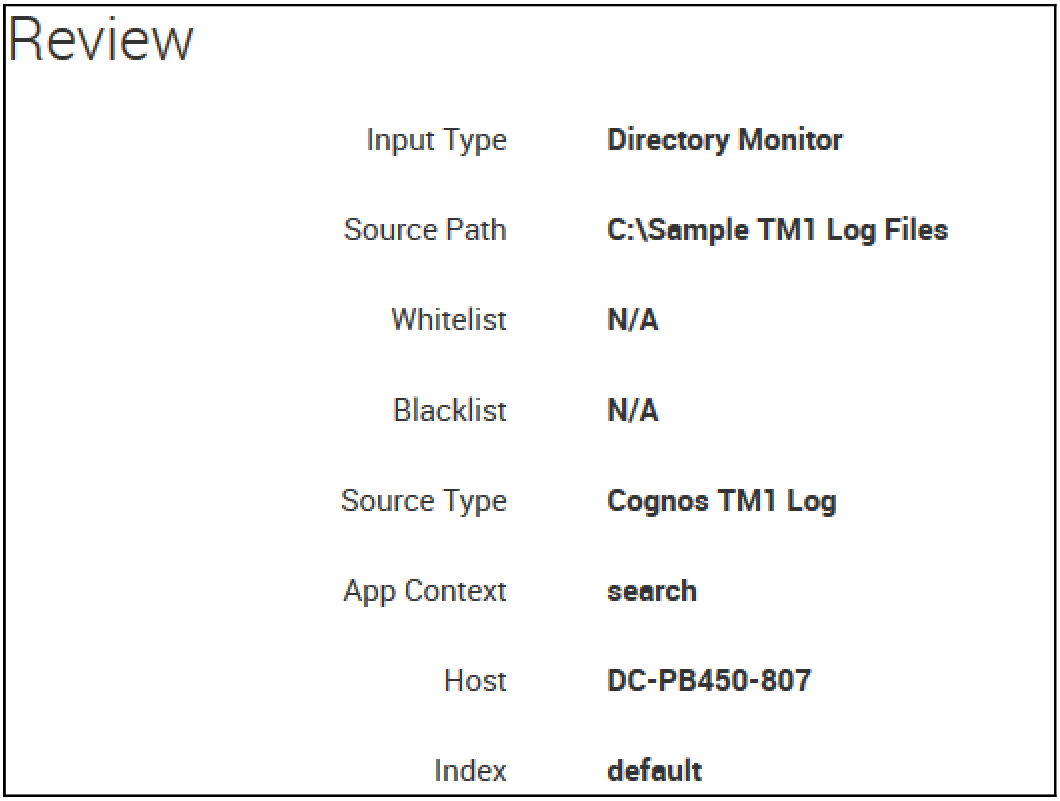
Source Type: Cognos TM1 Log

Source Type Category: Custom

Source Type Description: Application Log Files from Cognos TM1

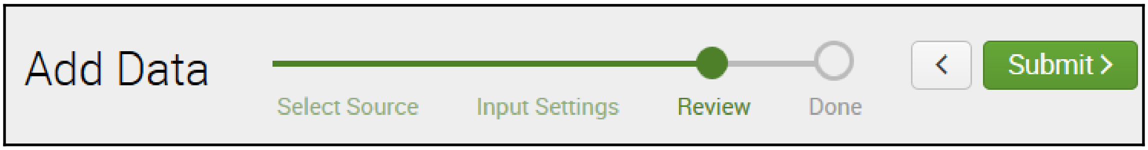

Automatic
Select
New
Cognos TM1 Log
filter
Application
Custom
Database
Email
Miscellaneous
Network & Security
Operating System
Structured
Web
✓ Cognos TM1 Log
Application Log Files from Cognos TM1
App C
ession on path
Segment in path
Host fielc

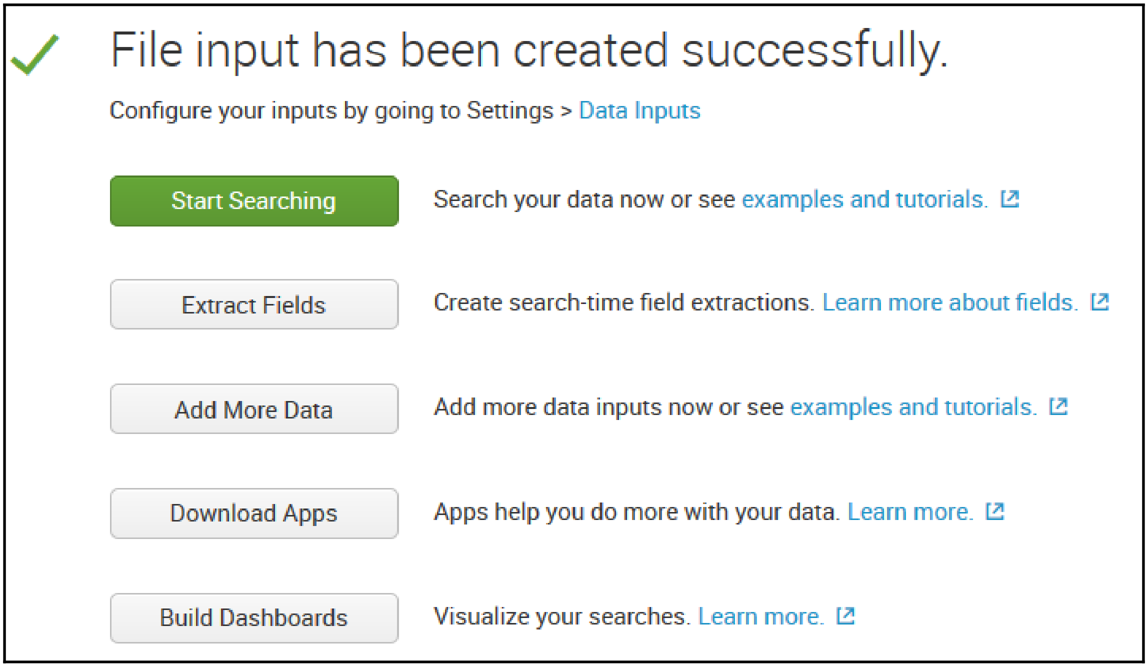

Add Data
Select Source
Input Settings
Review
Done
<
Review >



Review

Input Type	**Directory Monitor**
Source Path	**C:\Sample TM1 Log Files**
Whitelist	**N/A**
Blacklist	**N/A**
Source Type	**Cognos TM1 Log**
App Context	**search**
Host	**DC-PB450-807**
Index	**default**





✓ File input has been created successfully.

Configure your inputs by going to Settings > Data Inputs

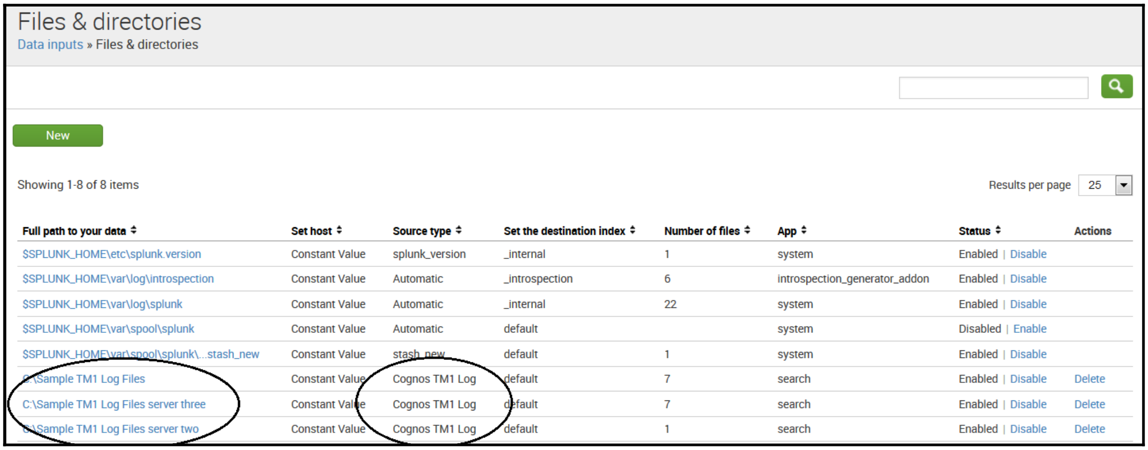
Search your data now or see examples and tutorials.

Create search-time field extractions. Learn more about fields.

Add more data inputs now or see examples and tutorials.

Apps help you do more with your data. Learn more.

Visualize your searches. Learn more.



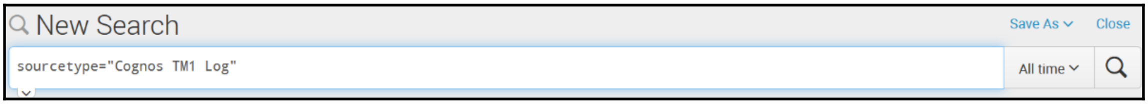
Files & directories

Data inputs » Files & directories

Showing 1-8 of 8 items

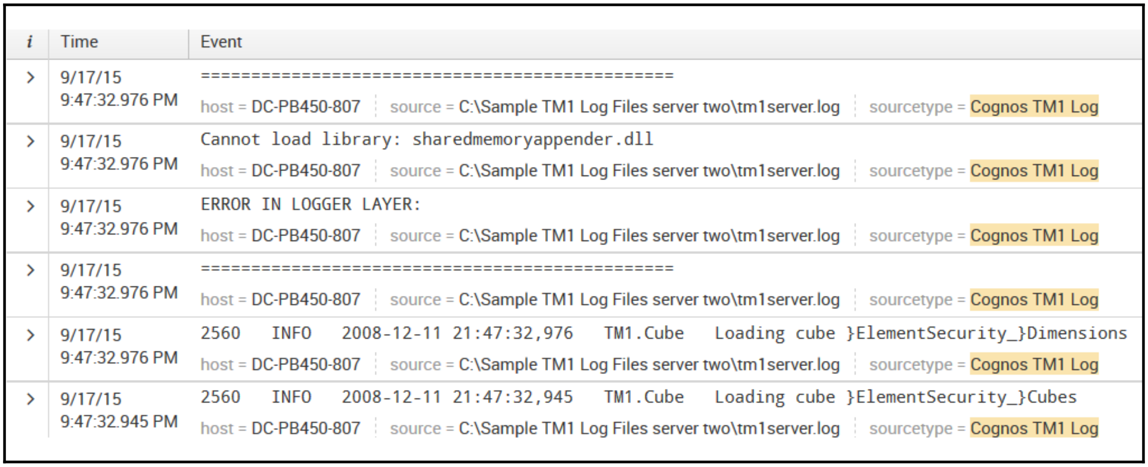
Results per page 25

Full path to your data	Set host	Source type	Set the destination index	Number of files	App	Status	Actions
$SPLUNK_HOME\etc\splunk.version	Constant Value	splunk_version	_internal	1	system	Enabled | Disable	
$SPLUNK_HOME\var\log\introspection	Constant Value	Automatic	_introspection	6	introspection_generator_addon	Enabled | Disable	
$SPLUNK_HOME\var\log\splunk	Constant Value	Automatic	_internal	22	system	Enabled | Disable	
$SPLUNK_HOME\var\spool\splunk	Constant Value	Automatic	default		system	Disabled | Enable	
$SPLUNK_HOME\var\spool\splunk\...stash_new	Constant Value	stash_new	default	1	system	Enabled | Disable	
C:\Sample TM1 Log Files	Constant Value	Cognos TM1 Log	default	7	search	Enabled | Disable	Delete
C:\Sample TM1 Log Files server three	Constant Value	Cognos TM1 Log	default	7	search	Enabled | Disable	Delete
C:\Sample TM1 Log Files server two	Constant Value	Cognos TM1 Log	default	1	search	Enabled | Disable	Delete

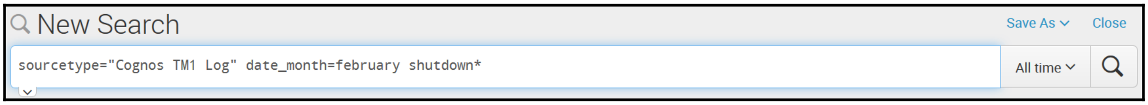


New Search

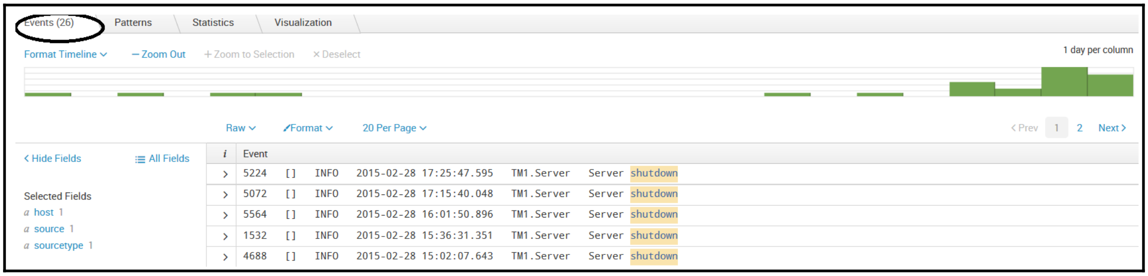
```
sourcetype="Cognos TM1 Log"
```



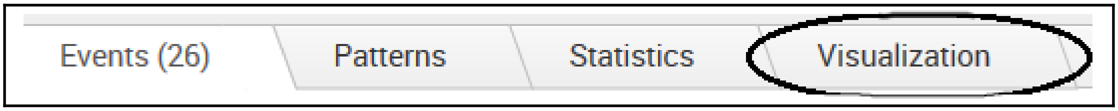
i	Time	Event
>	9/17/15 9:47:32.976 PM	== host = DC-PB450-807 ¦ source = C:\Sample TM1 Log Files server two\tm1server.log ¦ sourcetype = Cognos TM1 Log
>	9/17/15 9:47:32.976 PM	Cannot load library: sharedmemoryappender.dll host = DC-PB450-807 ¦ source = C:\Sample TM1 Log Files server two\tm1server.log ¦ sourcetype = Cognos TM1 Log
>	9/17/15 9:47:32.976 PM	ERROR IN LOGGER LAYER: host = DC-PB450-807 ¦ source = C:\Sample TM1 Log Files server two\tm1server.log ¦ sourcetype = Cognos TM1 Log
>	9/17/15 9:47:32.976 PM	== host = DC-PB450-807 ¦ source = C:\Sample TM1 Log Files server two\tm1server.log ¦ sourcetype = Cognos TM1 Log
>	9/17/15 9:47:32.976 PM	2560 INFO 2008-12-11 21:47:32,976 TM1.Cube Loading cube }ElementSecurity_}Dimensions host = DC-PB450-807 ¦ source = C:\Sample TM1 Log Files server two\tm1server.log ¦ sourcetype = Cognos TM1 Log
>	9/17/15 9:47:32.945 PM	2560 INFO 2008-12-11 21:47:32,945 TM1.Cube Loading cube }ElementSecurity_}Cubes host = DC-PB450-807 ¦ source = C:\Sample TM1 Log Files server two\tm1server.log ¦ sourcetype = Cognos TM1 Log

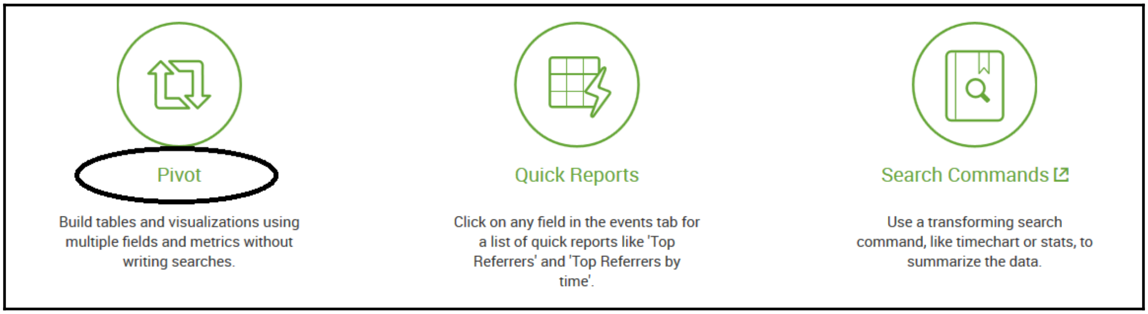


```
sourcetype="Cognos TM1 Log" date_month=february shutdown*
```



```
5224  []  INFO  2015-02-28 17:25:47.595  TM1.Server  Server shutdown
5072  []  INFO  2015-02-28 17:15:40.048  TM1.Server  Server shutdown
5564  []  INFO  2015-02-28 16:01:50.896  TM1.Server  Server shutdown
1532  []  INFO  2015-02-28 15:36:31.351  TM1.Server  Server shutdown
4688  []  INFO  2015-02-28 15:02:07.643  TM1.Server  Server shutdown
```





Pivot

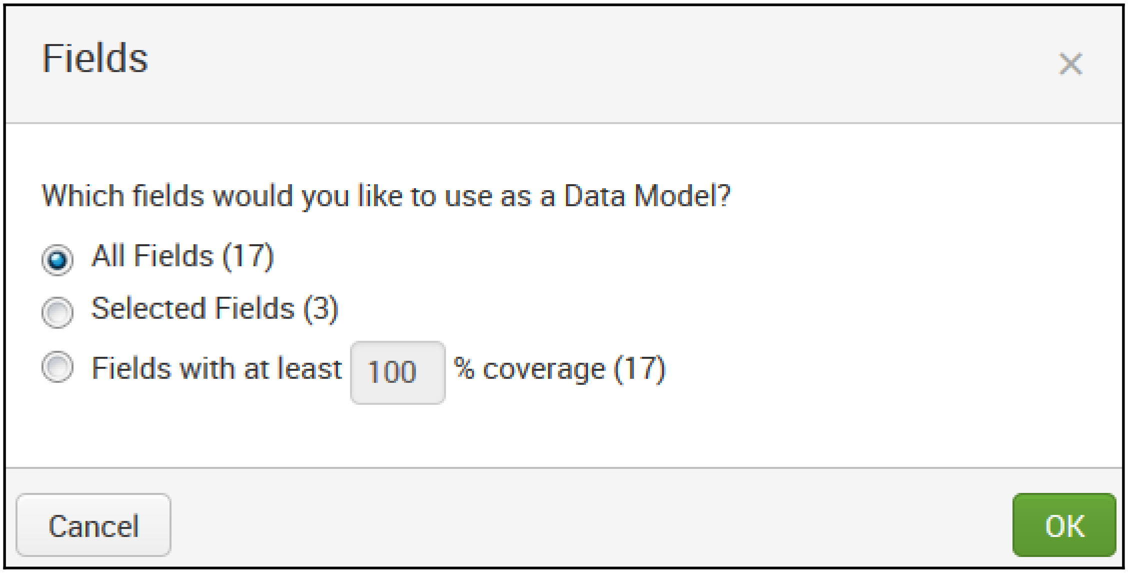
Build tables and visualizations using multiple fields and metrics without writing searches.

Quick Reports

Click on any field in the events tab for a list of quick reports like 'Top Referrers' and 'Top Referrers by time'.

Search Commands

Use a transforming search command, like timechart or stats, to summarize the data.



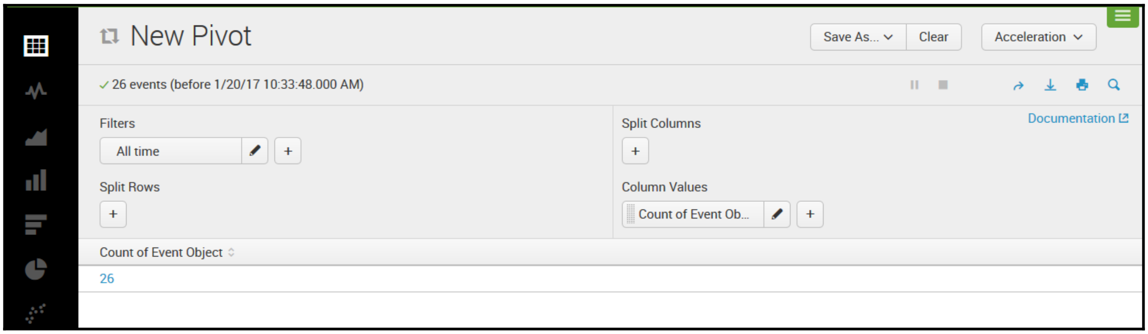
Fields

Which fields would you like to use as a Data Model?

● All Fields (17)

○ Selected Fields (3)

○ Fields with at least 100 % coverage (17)



New Pivot

✓ 26 events (before 1/20/17 10:33:48.000 AM)

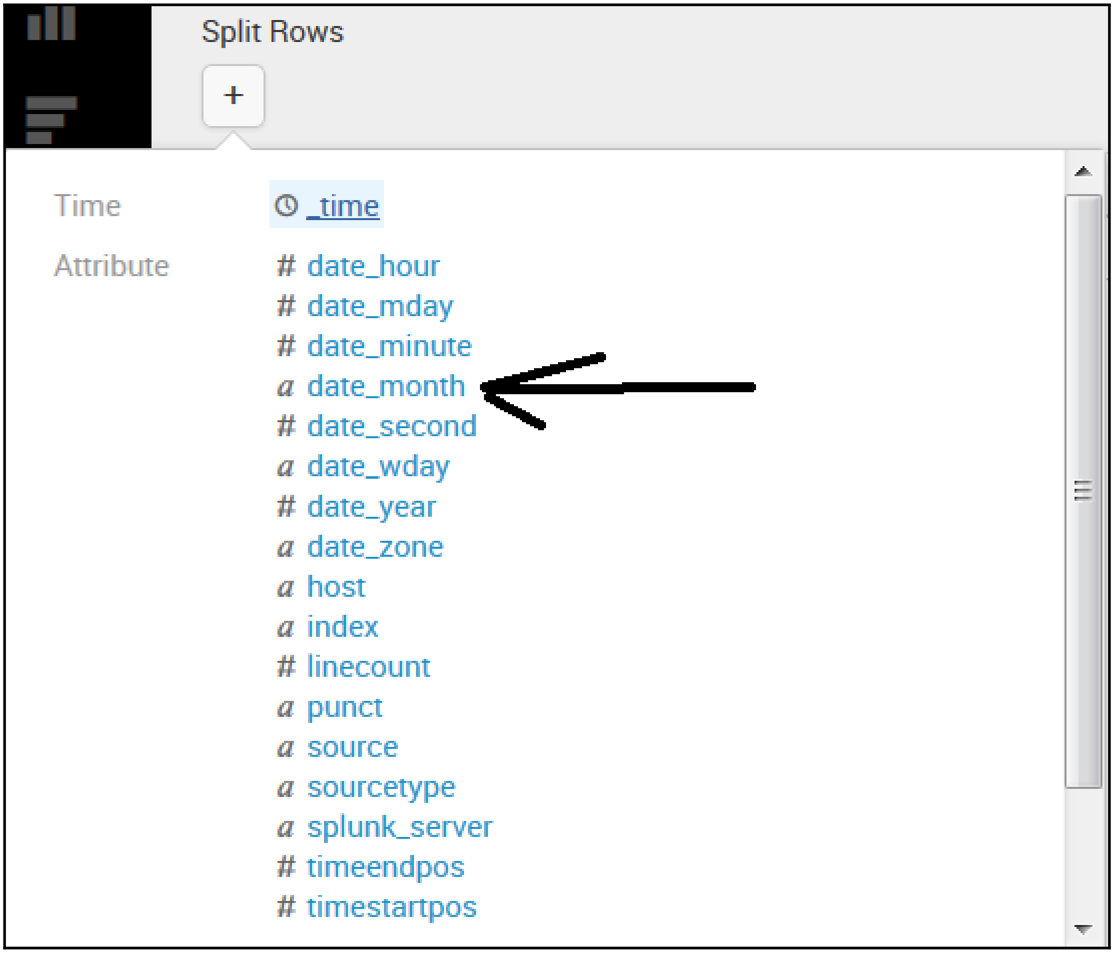
Filters: All time

Split Rows

Split Columns

Column Values: Count of Event Ob...

Count of Event Object
26

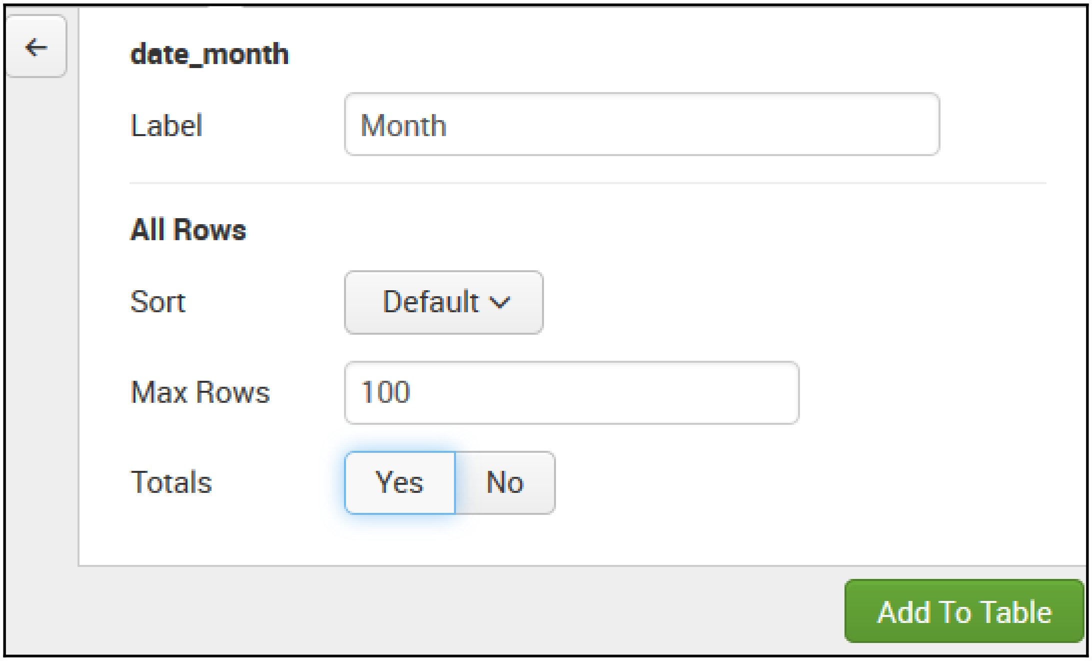


Split Rows

+

Time ◷ _time

Attribute
# date_hour
# date_mday
# date_minute
a date_month
# date_second
a date_wday
# date_year
a date_zone
a host
a index
# linecount
a punct
a source
a sourcetype
a splunk_server
# timeendpos
# timestartpos



date_month

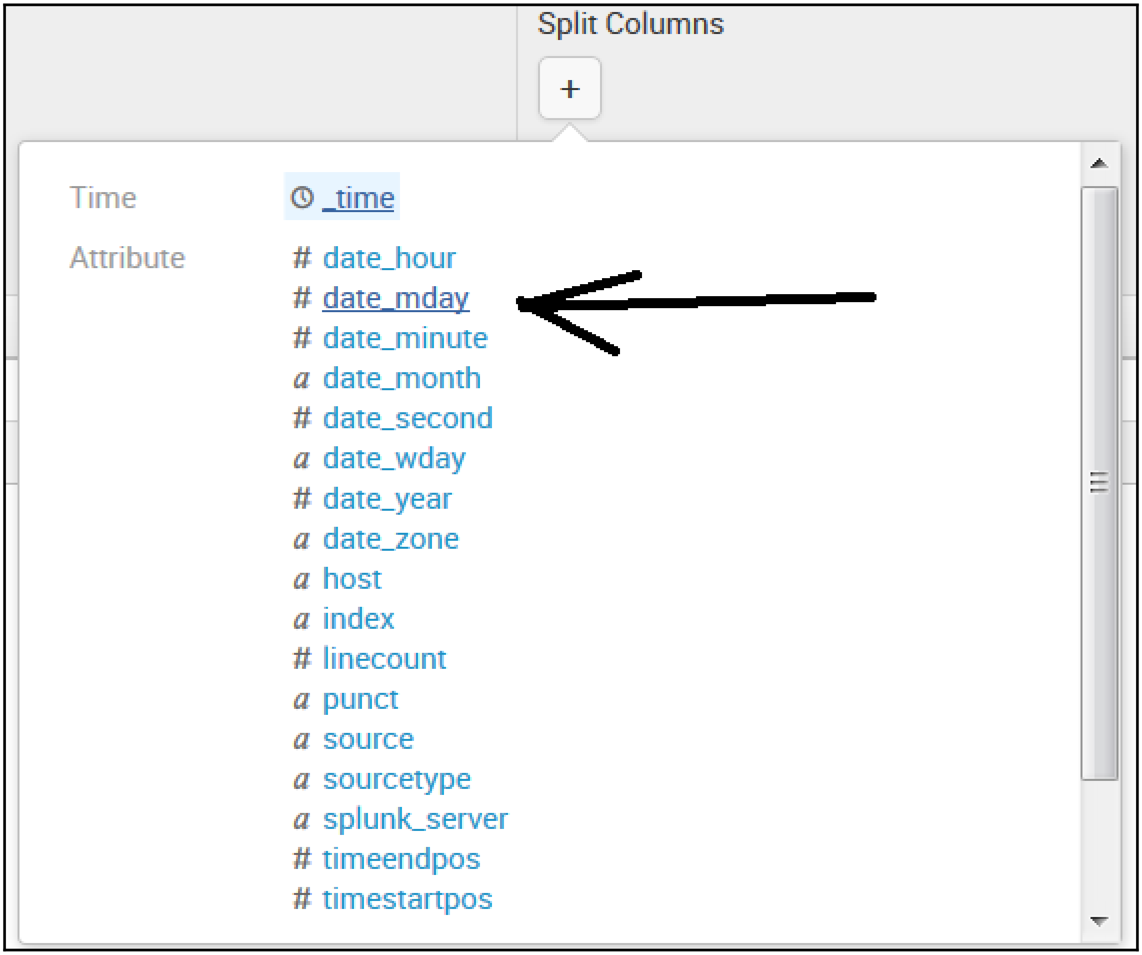
Label: Month

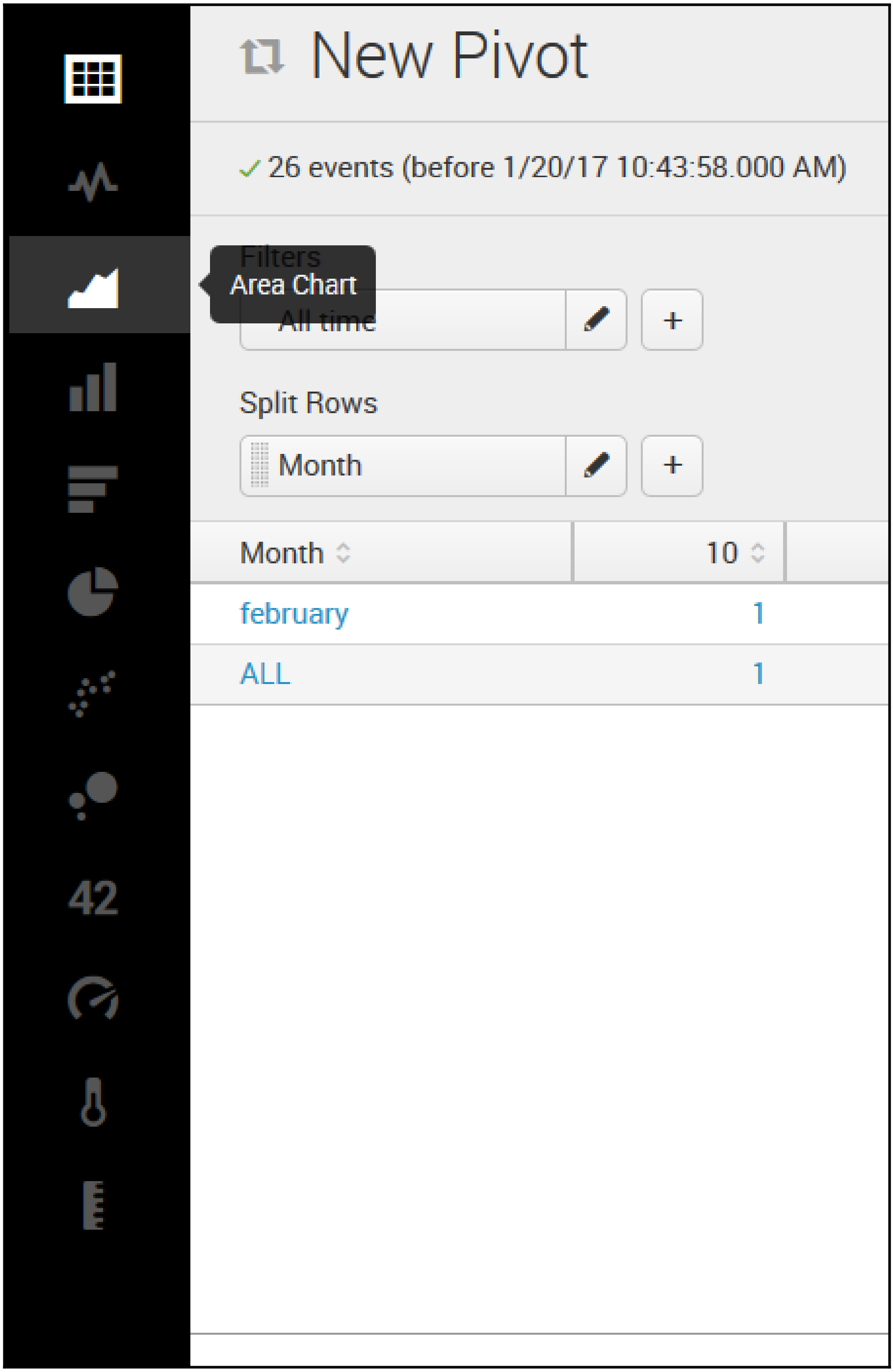
All Rows

Sort: Default

Max Rows: 100

Totals: [✓ Yes] No


Split Columns
+
Time
_time
Attribute
# date_hour
# date_mday
# date_minute
a date_month
# date_second
a date_wday
# date_year
a date_zone
a host
a index
# linecount
a punct
a source
a sourcetype
a splunk_server
# timeendpos
# timestartpos



New Pivot

✓ 26 events (before 1/20/17 10:43:58.000 AM)

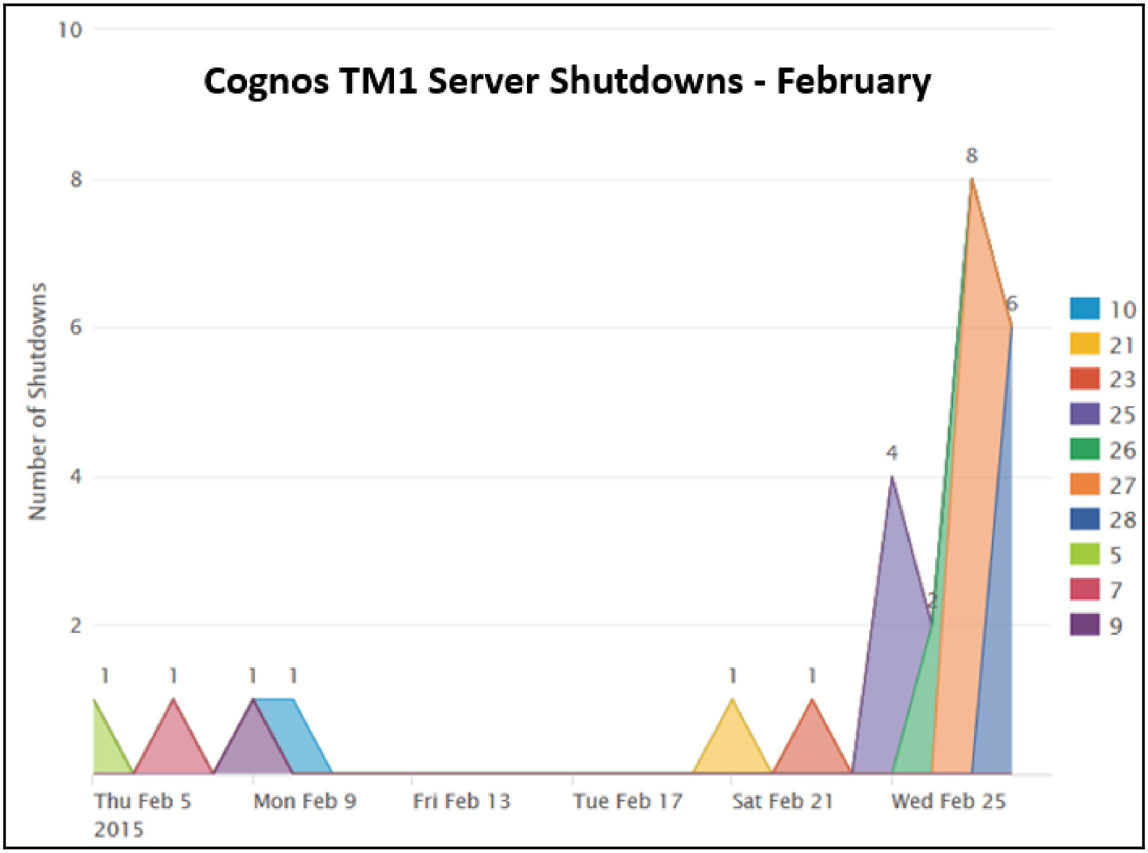
Filters

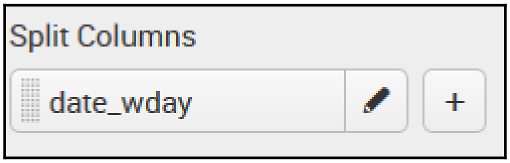
All time

Split Rows

Month

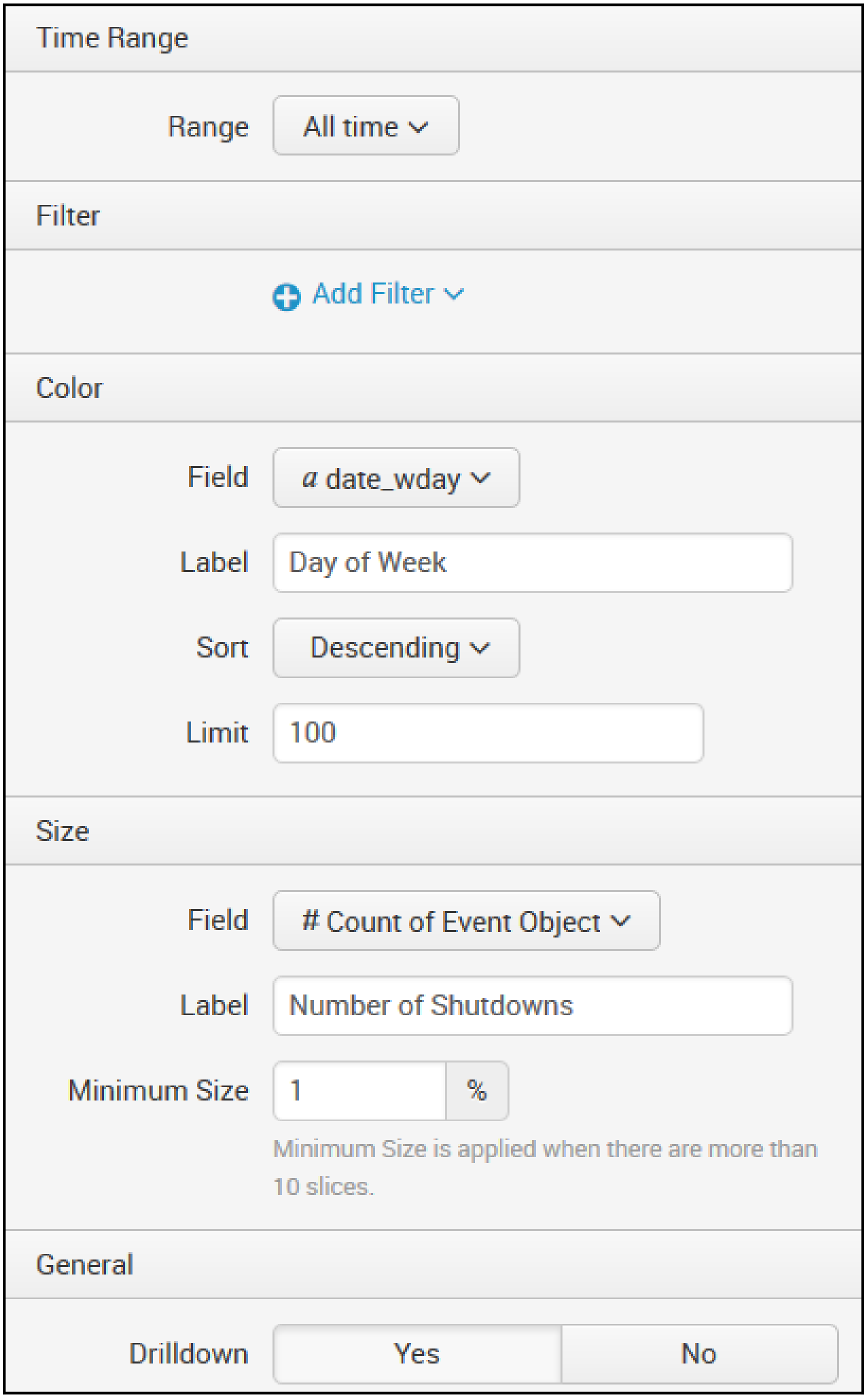
Month	10
february	1
ALL	1





Split Columns

date_wday



Time Range

Range: All time

Filter

Add Filter

Color

Field: a date_wday

Label: Day of Week

Sort: Descending

Limit: 100

Size

Field: # Count of Event Object

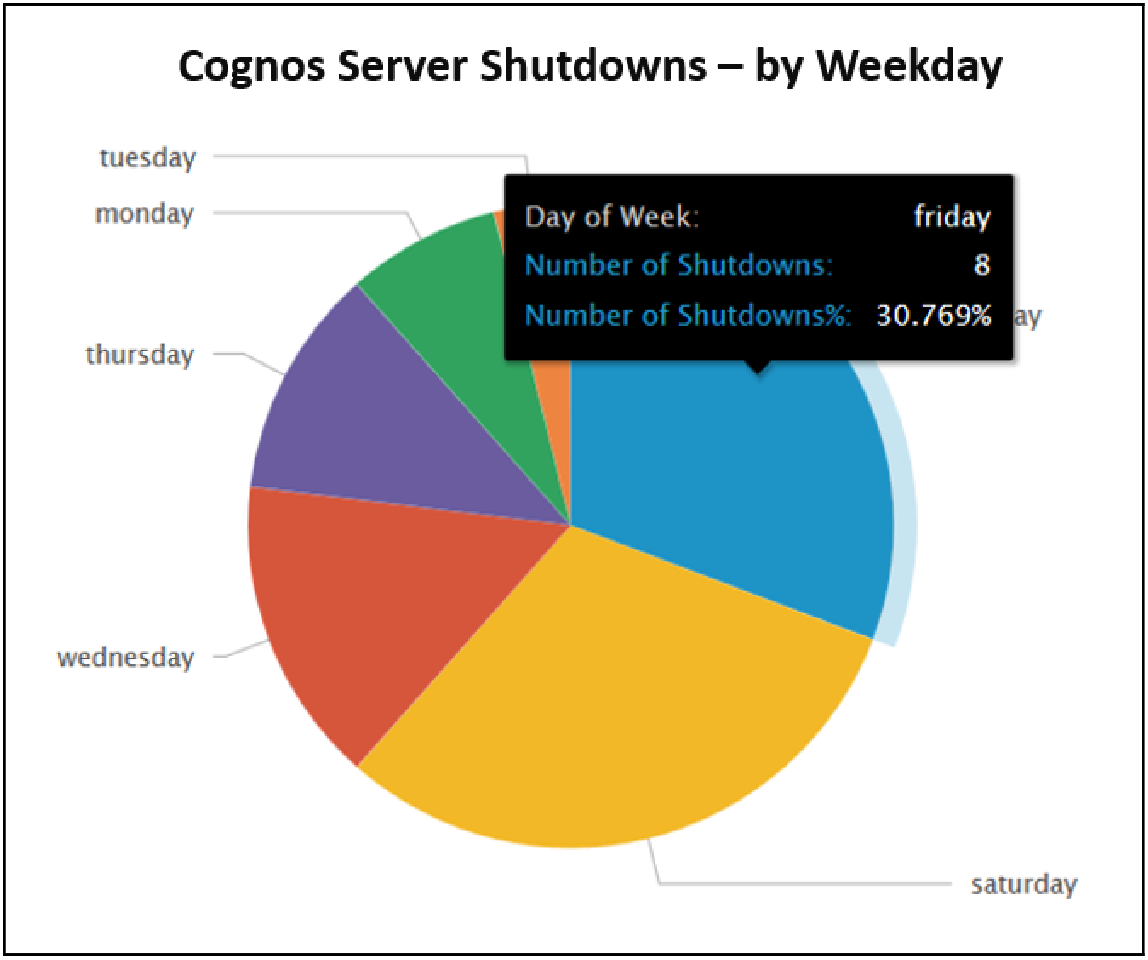
Label: Number of Shutdowns

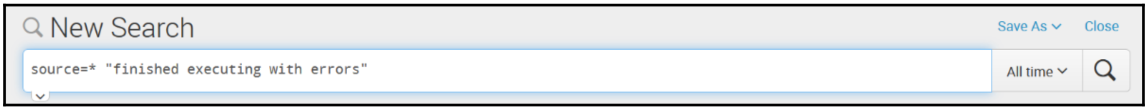
Minimum Size: 1 %

Minimum Size is applied when there are more than 10 slices.

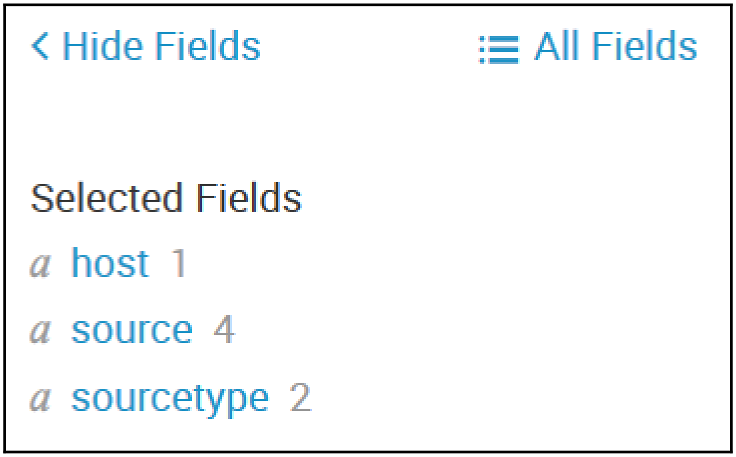
General

Drilldown: Yes No





```
source=* "finished executing with errors"
```



< Hide Fields

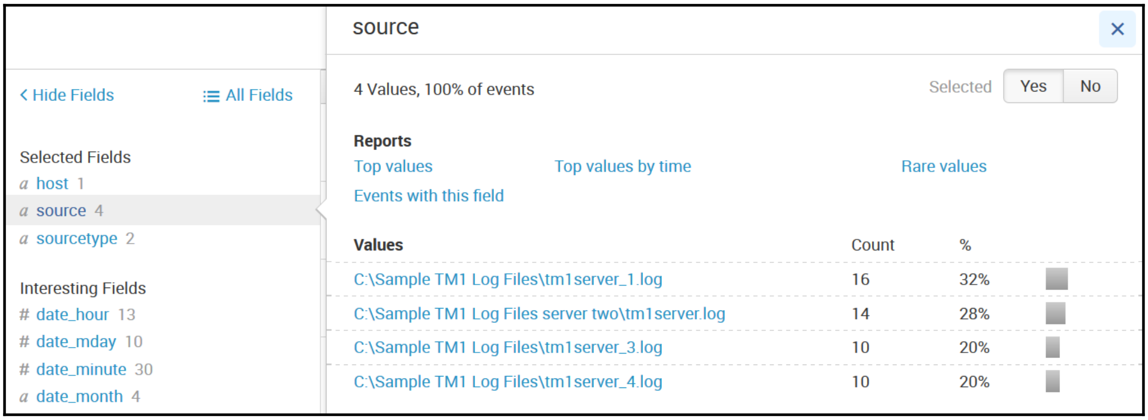
☰ All Fields

Selected Fields

a host 1

a source 4

a sourcetype 2



< Hide Fields

☰ All Fields

Selected Fields

a host 1

a source 4

a sourcetype 2

Interesting Fields

date_hour 13

date_mday 10

date_minute 30

a date_month 4

source ✕

4 Values, 100% of events

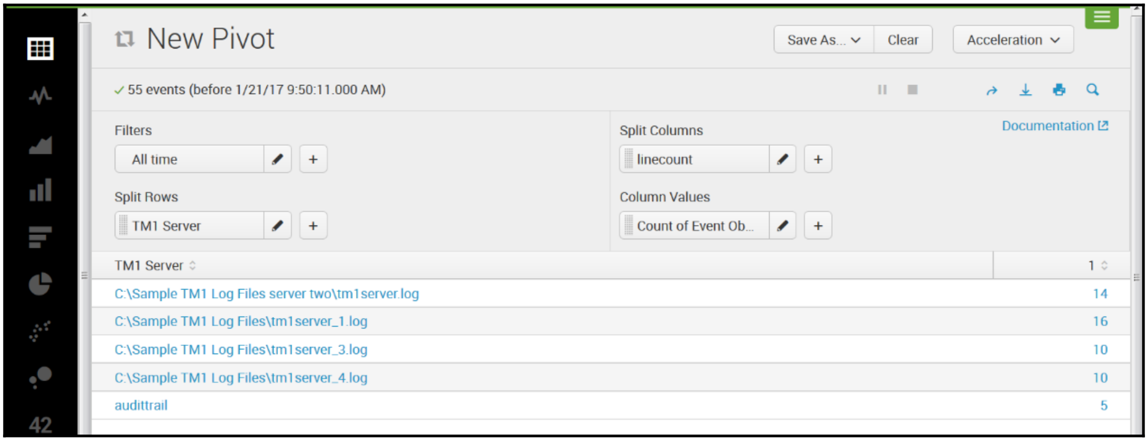
Selected Yes No

Reports

Top values | Top values by time | Rare values

Events with this field

Values	Count	%
C:\Sample TM1 Log Files\tm1server_1.log	16	32%
C:\Sample TM1 Log Files server two\tm1server.log	14	28%
C:\Sample TM1 Log Files\tm1server_3.log	10	20%
C:\Sample TM1 Log Files\tm1server_4.log	10	20%



New Pivot

Save As... | Clear | Acceleration

✓ 55 events (before 1/21/17 9:50:11.000 AM)

Documentation

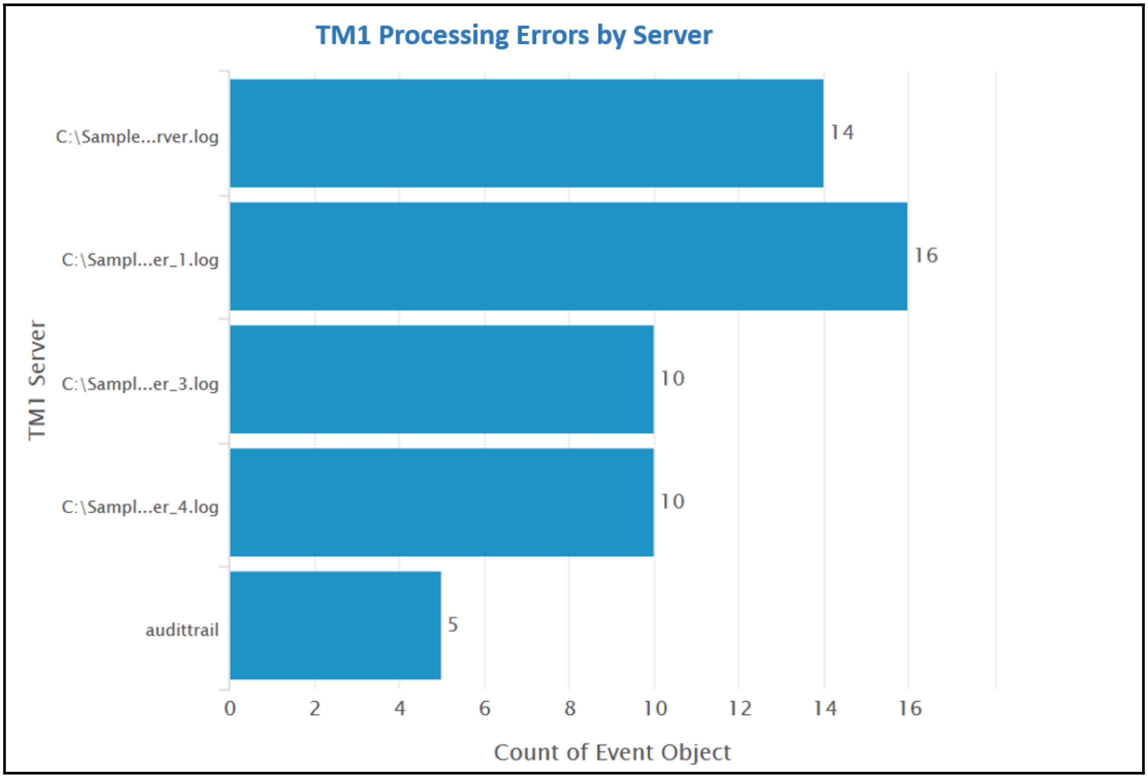
Filters: All time

Split Rows: TM1 Server

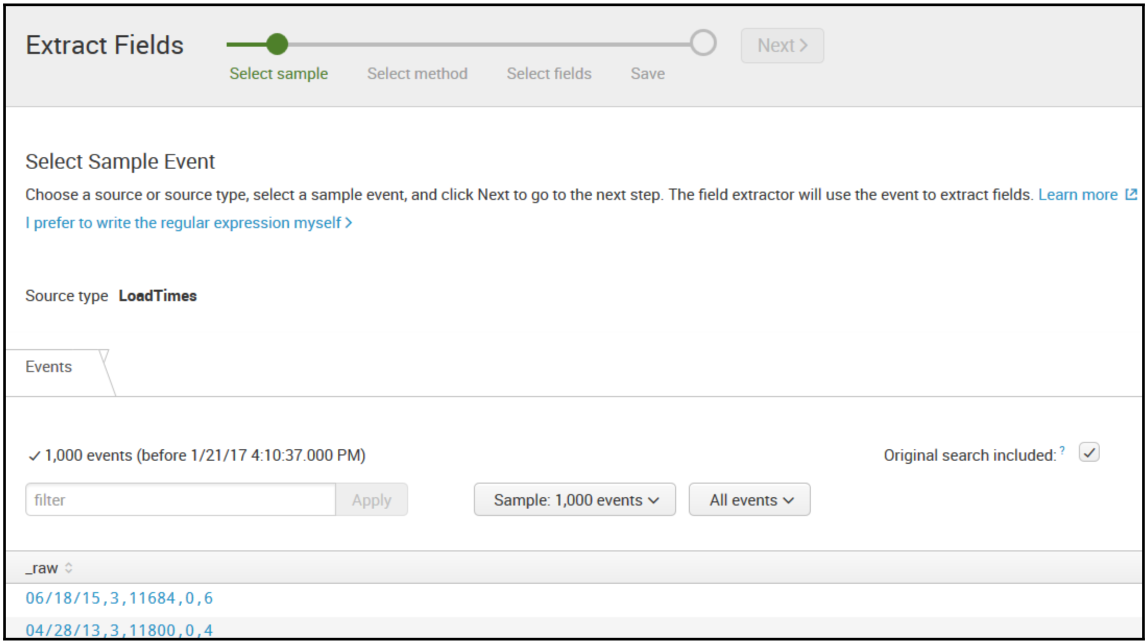
Split Columns: linecount

Column Values: Count of Event Ob...

TM1 Server	1
C:\Sample TM1 Log Files server two\tm1server.log	14
C:\Sample TM1 Log Files\tm1server_1.log	16
C:\Sample TM1 Log Files\tm1server_3.log	10
C:\Sample TM1 Log Files\tm1server_4.log	10
audittrail	5


TM1 Processing Errors by Server
TM1 Server
C:\Sample...rver.log
14
C:\Sampl...er_1.log
16
C:\Sampl...er_3.log
10
C:\Sampl...er_4.log
10
audittrail
5
0
2
4
6
8
10
12
14
16
Count of Event Object

⊕ Extract New Fields



Select Sample Event

Choose a source or source type, select a sample event, and click Next to go to the next step. The field extractor will use the event to extract fields. Learn more

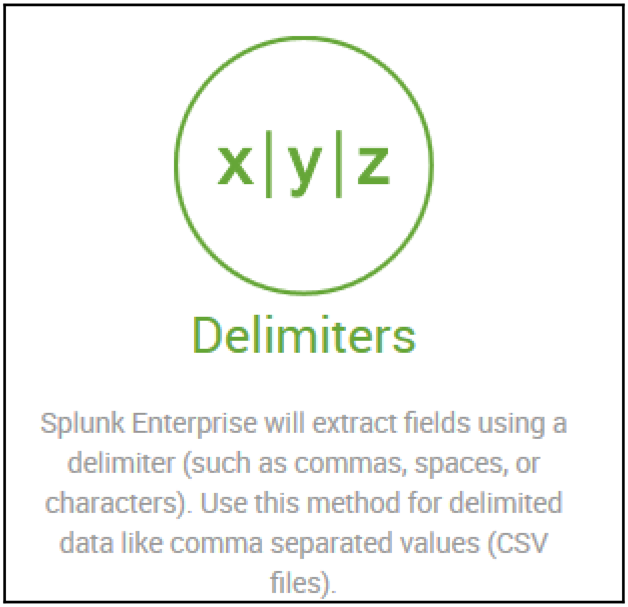
I prefer to write the regular expression myself >

Source type **LoadTimes**

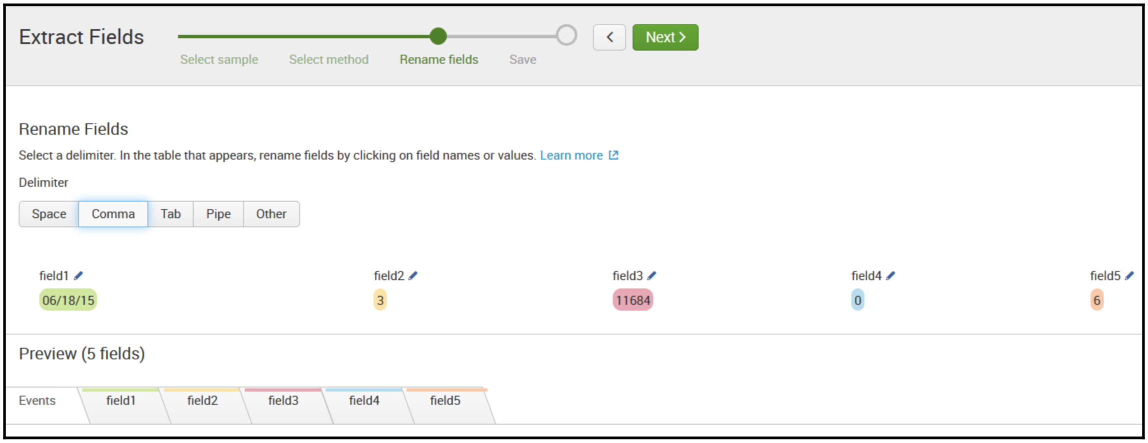
✓ 1,000 events (before 1/21/17 4:10:37.000 PM)

Original search included: ☑

```
06/18/15,3,11684,0,6
04/28/13,3,11800,0,4
```



Splunk Enterprise will extract fields using a delimiter (such as commas, spaces, or characters). Use this method for delimited data like comma separated values (CSV files).



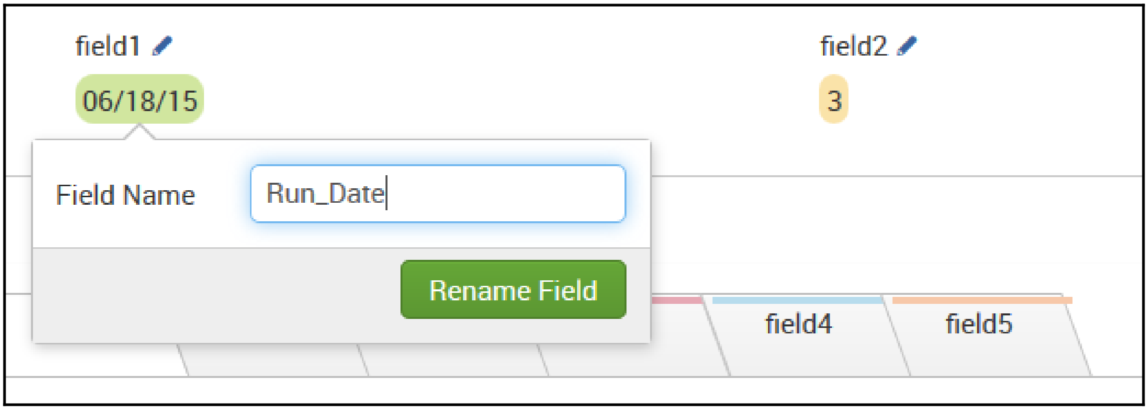
Rename Fields

Select a delimiter. In the table that appears, rename fields by clicking on field names or values. Learn more

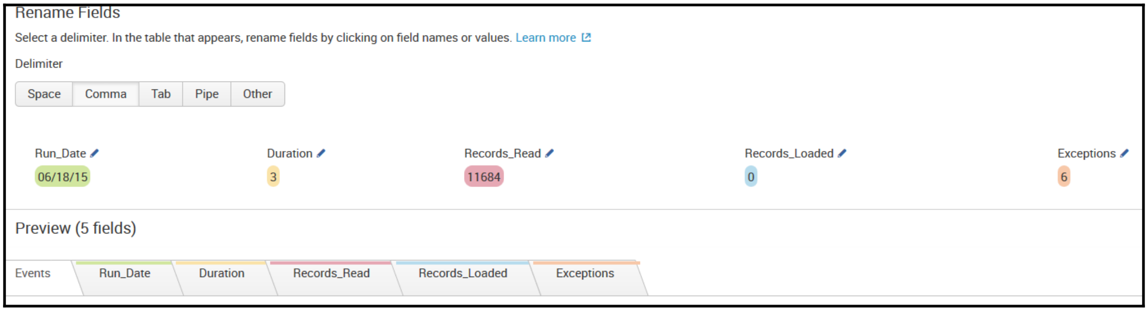
Delimiter: Space [✓ Comma] Tab Pipe Other

field1	field2	field3	field4	field5
06/18/15	3	11684	0	6

Preview (5 fields)



Field Name: Run_Date

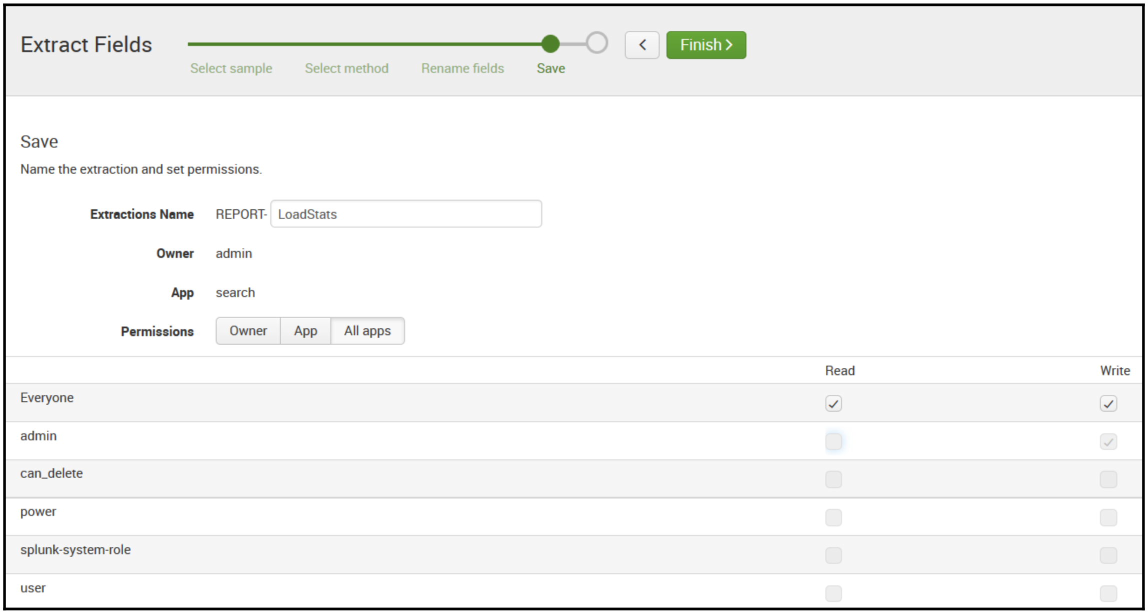

Rename Fields

Select a delimiter. In the table that appears, rename fields by clicking on field names or values. Learn more

Delimiter

Run_Date	Duration	Records_Read	Records_Loaded	Exceptions
06/18/15	3	11684	0	6

Preview (5 fields)



Save

Name the extraction and set permissions.

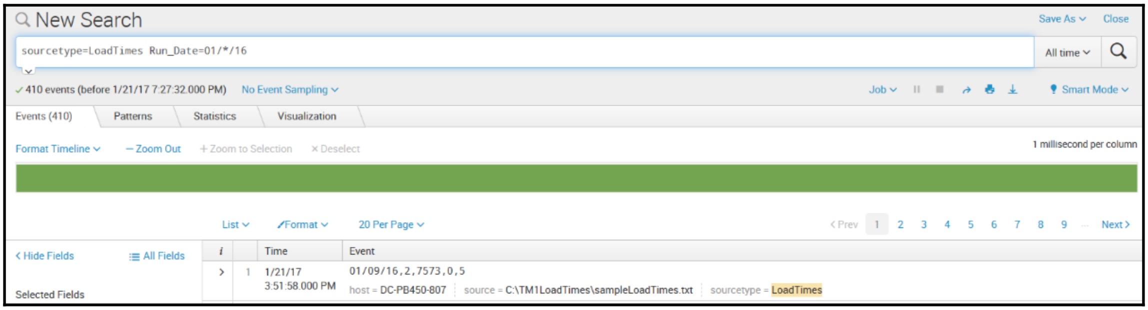
Extractions Name REPORT- LoadStats

Owner admin

App search

Permissions Owner | App | [✓ All apps]

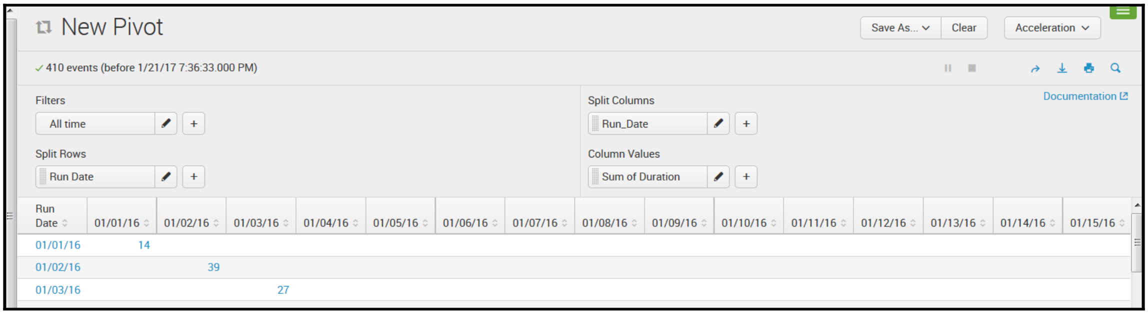
	Read	Write
Everyone	☑	☑
admin	□	☑
can_delete	□	□
power	□	□
splunk-system-role	□	□
user	□	□



```
sourcetype=LoadTimes Run_Date=01/*/16
```

✓ 410 events (before 1/21/17 7:27:32.000 PM)

	i	Time	Event
>	1	1/21/17 3:51:58.000 PM	01/09/16,2,7573,0,5 host = DC-PB450-807 source = C:\TM1LoadTimes\sampleLoadTimes.txt sourcetype = LoadTimes



✓ 410 events (before 1/21/17 7:36:33.000 PM)

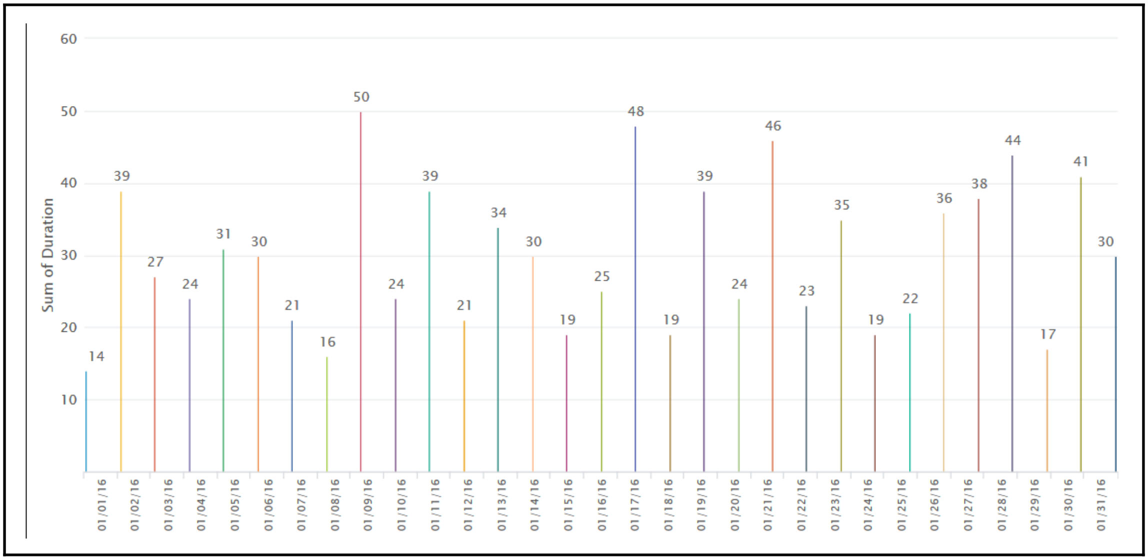
Filters: All time

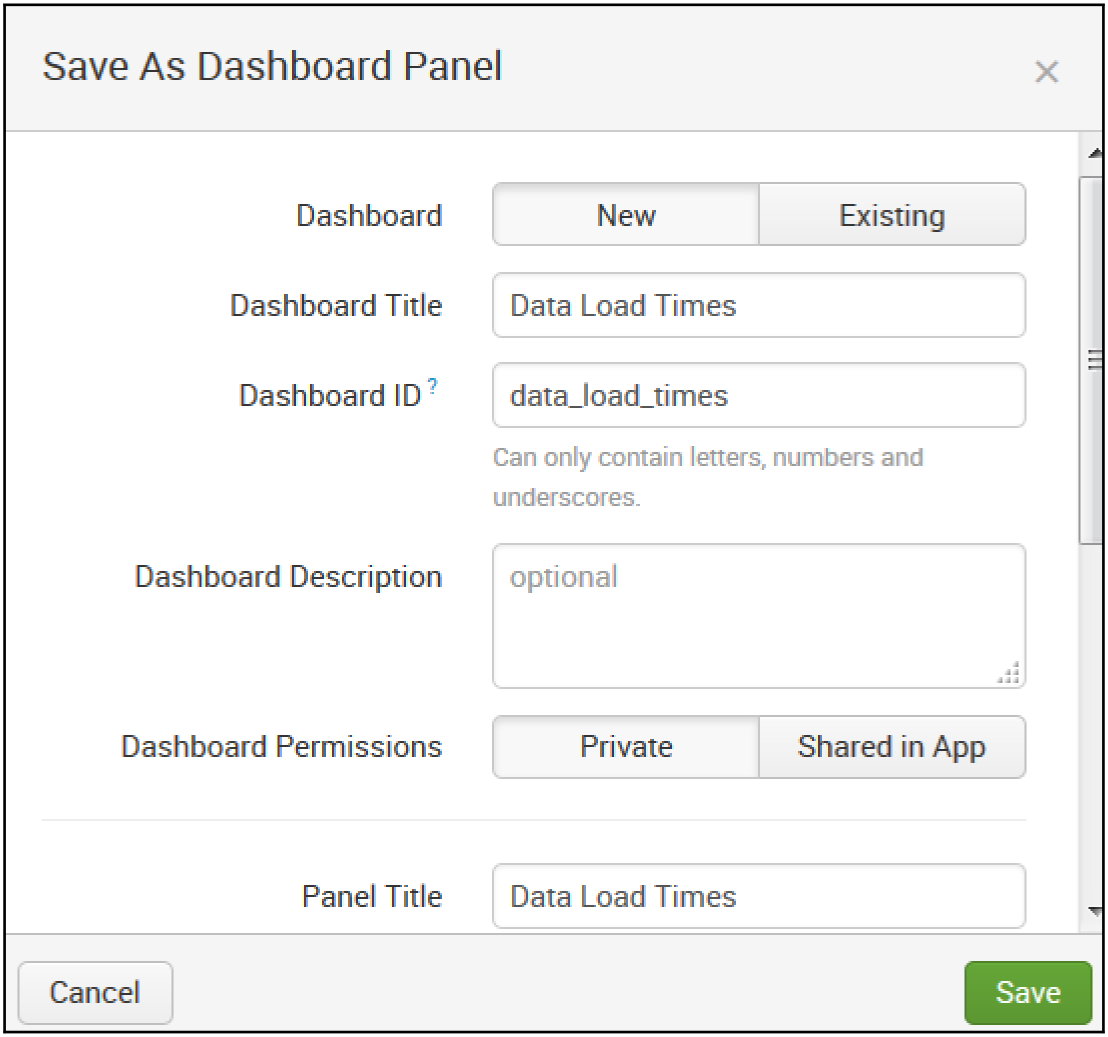
Split Rows: Run Date

Split Columns: Run_Date

Column Values: Sum of Duration

Run Date	01/01/16	01/02/16	01/03/16	01/04/16	01/05/16	01/06/16	01/07/16	01/08/16	01/09/16	01/10/16	01/11/16	01/12/16	01/13/16	01/14/16	01/15/16
01/01/16	14														
01/02/16		39													
01/03/16			27												


Sum of Duration
60
50
40
30
20
10
14
39
27
24
31
30
21
16
50
24
39
21
34
30
19
25
48
19
39
24
46
23
35
19
22
36
38
44
17
41
30
01/01/16
01/02/16
01/03/16
01/04/16
01/05/16
01/06/16
01/07/16
01/08/16
01/09/16
01/10/16
01/11/16
01/12/16
01/13/16
01/14/16
01/15/16
01/16/16
01/17/16
01/18/16
01/19/16
01/20/16
01/21/16
01/22/16
01/23/16
01/24/16
01/25/16
01/26/16
01/27/16
01/28/16
01/29/16
01/30/16
01/31/16



Save As Dashboard Panel

Dashboard: New | Existing

Dashboard Title: Data Load Times

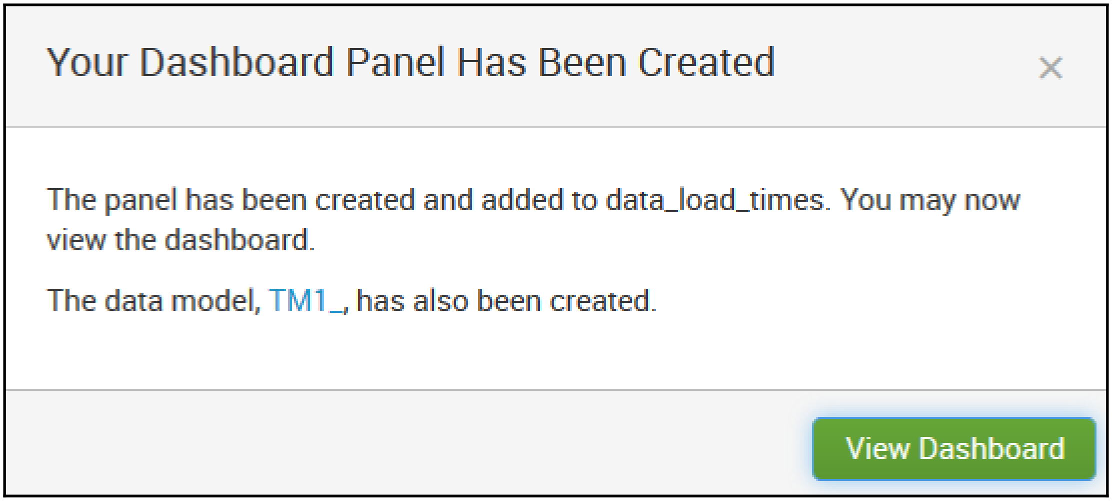
Dashboard ID ?: data_load_times

Can only contain letters, numbers and underscores.

Dashboard Description: optional

Dashboard Permissions: Private | Shared in App

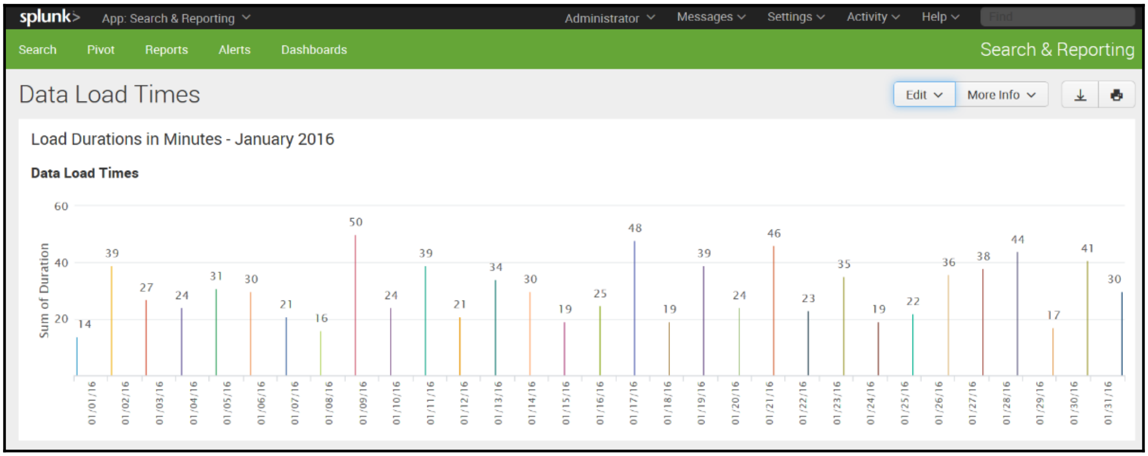
Panel Title: Data Load Times



Your Dashboard Panel Has Been Created

The panel has been created and added to data_load_times. You may now view the dashboard.

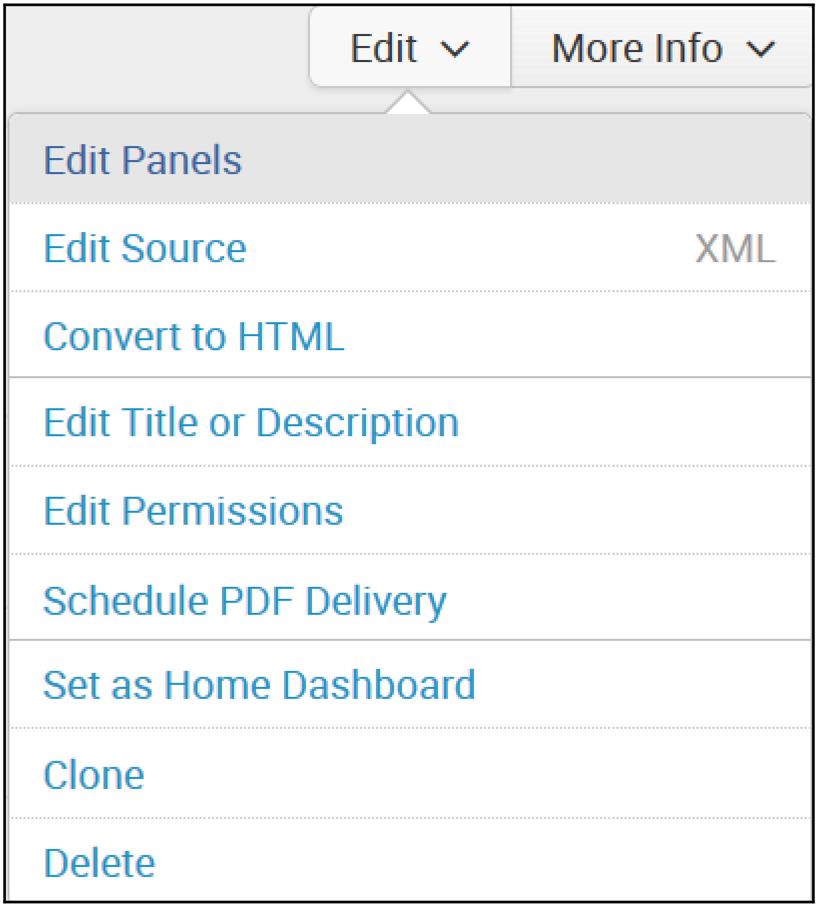
The data model, TM1_, has also been created.

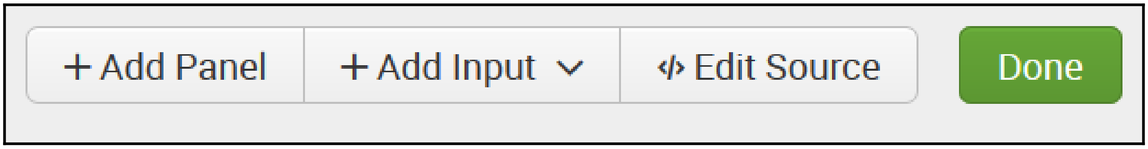


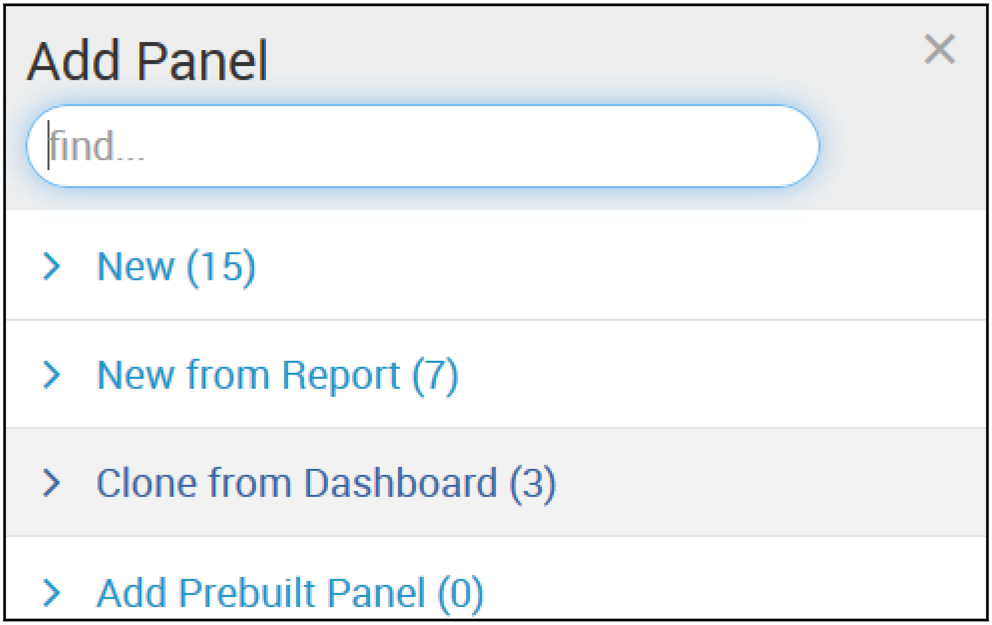
Data Load Times

Load Durations in Minutes - January 2016

Data Load Times


Edit
More Info
Edit Panels
Edit Source
XML
Convert to HTML
Edit Title or Description
Edit Permissions
Schedule PDF Delivery
Set as Home Dashboard
Clone
Delete


+ Add Panel
+ Add Input
Edit Source
Done



Add Panel

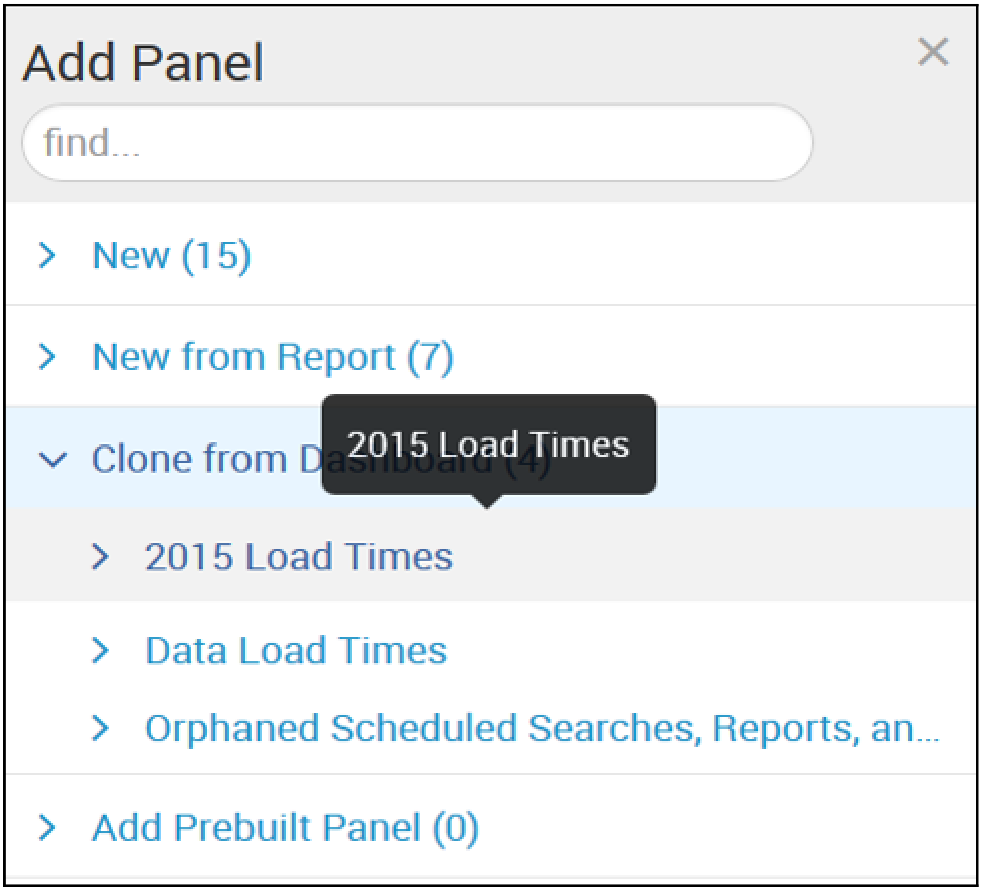
find...

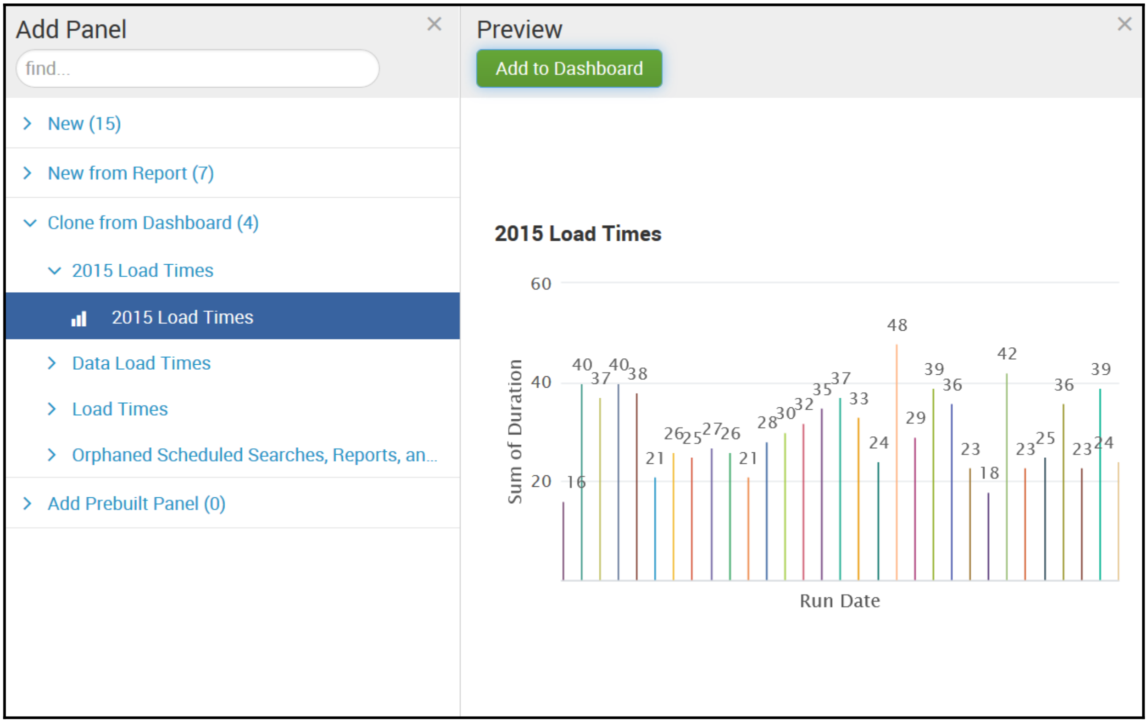
New (15)

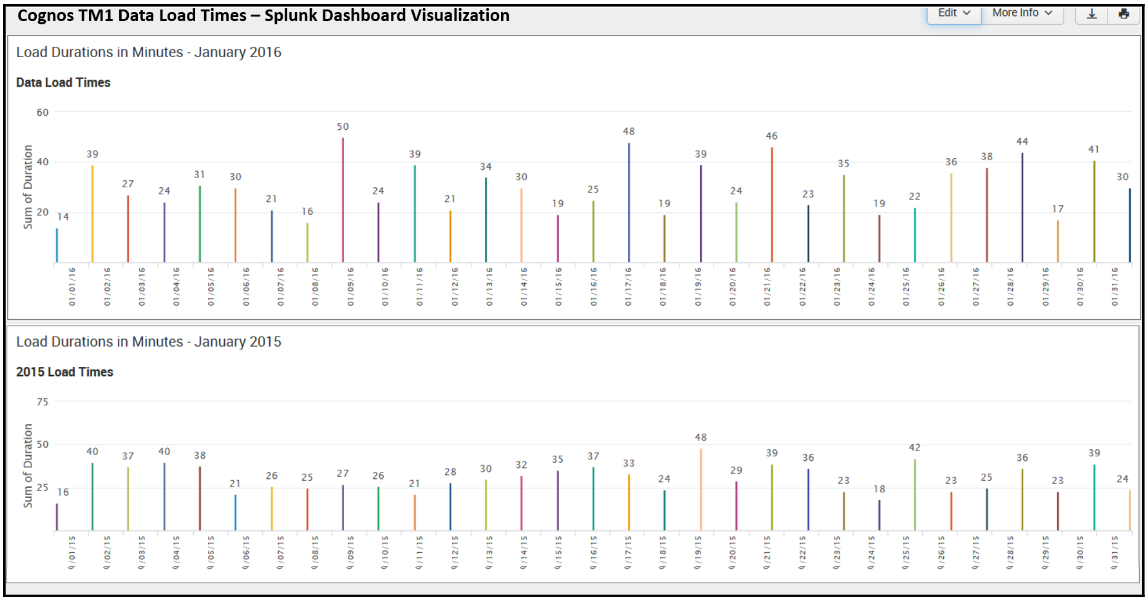
New from Report (7)

Clone from Dashboard (3)

Add Prebuilt Panel (0)


Add Panel
×
find...
New (15)
New from Report (7)
Clone from Dashboard (4)
2015 Load Times
2015 Load Times
Data Load Times
Orphaned Scheduled Searches, Reports, an...
Add Prebuilt Panel (0)




Cognos TM1 Data Load Times – Splunk Dashboard Visualization

Load Durations in Minutes - January 2016

Data Load Times

Load Durations in Minutes - January 2015

2015 Load Times