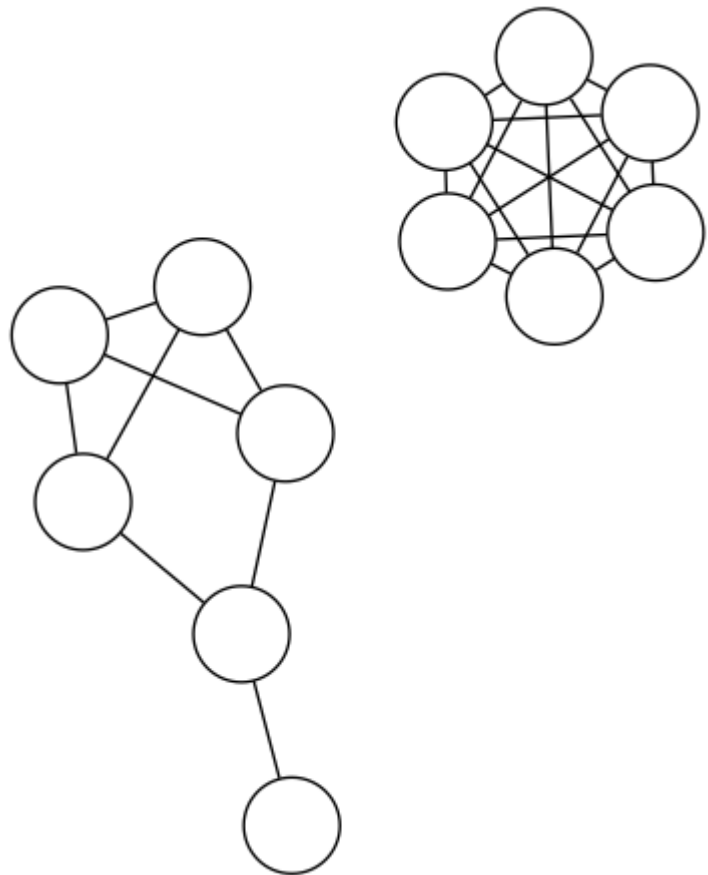
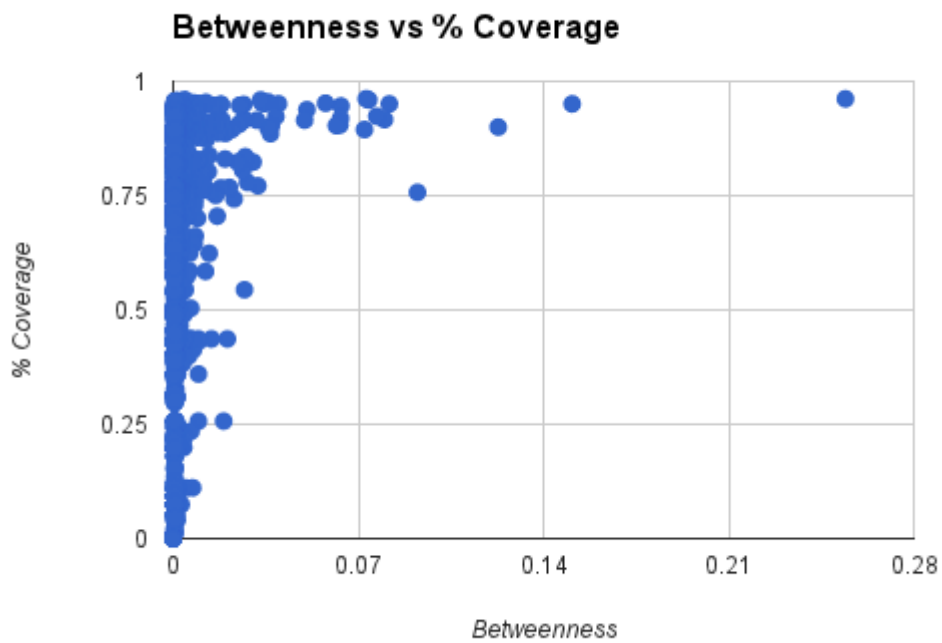
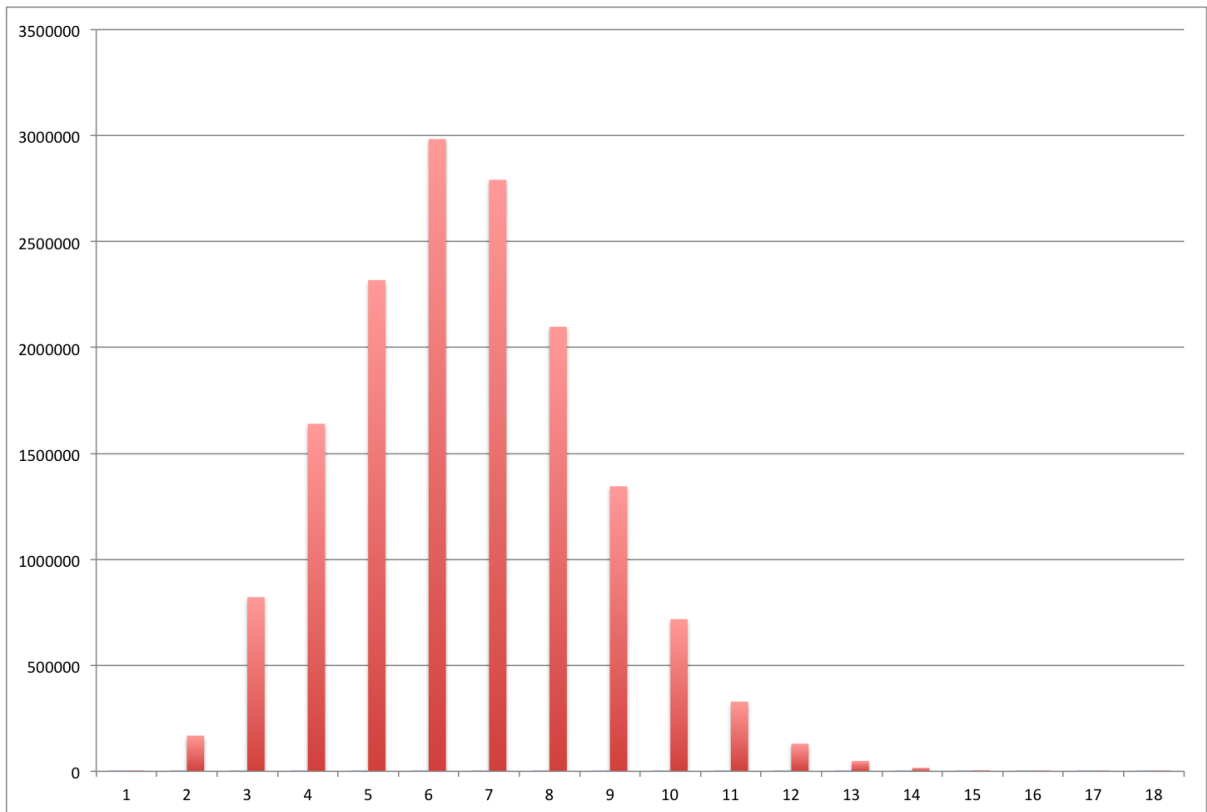
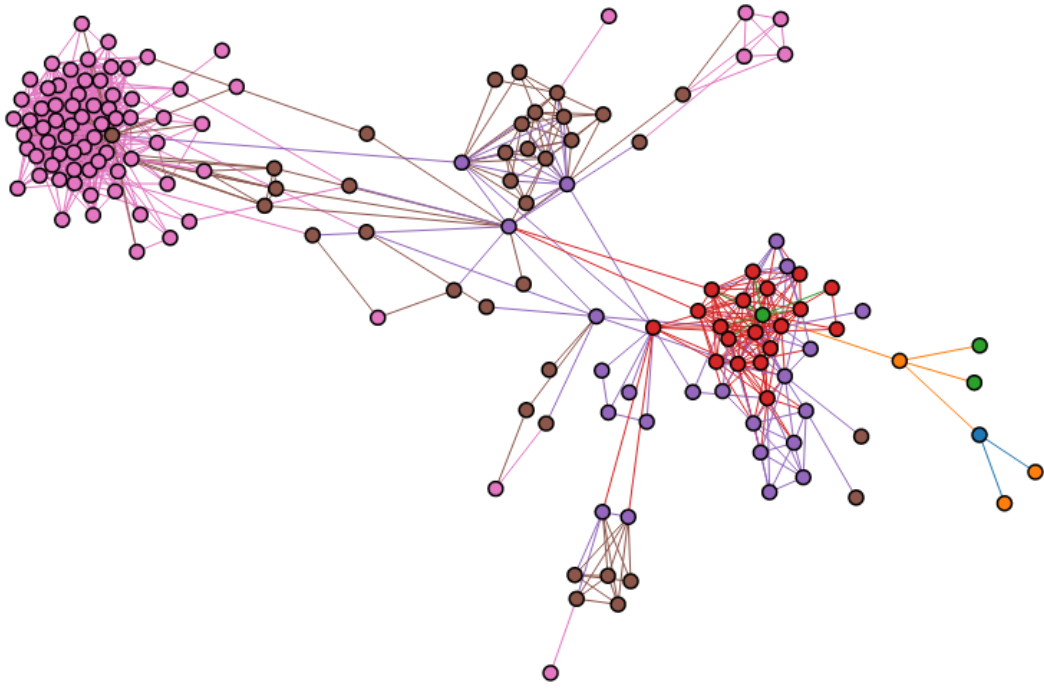


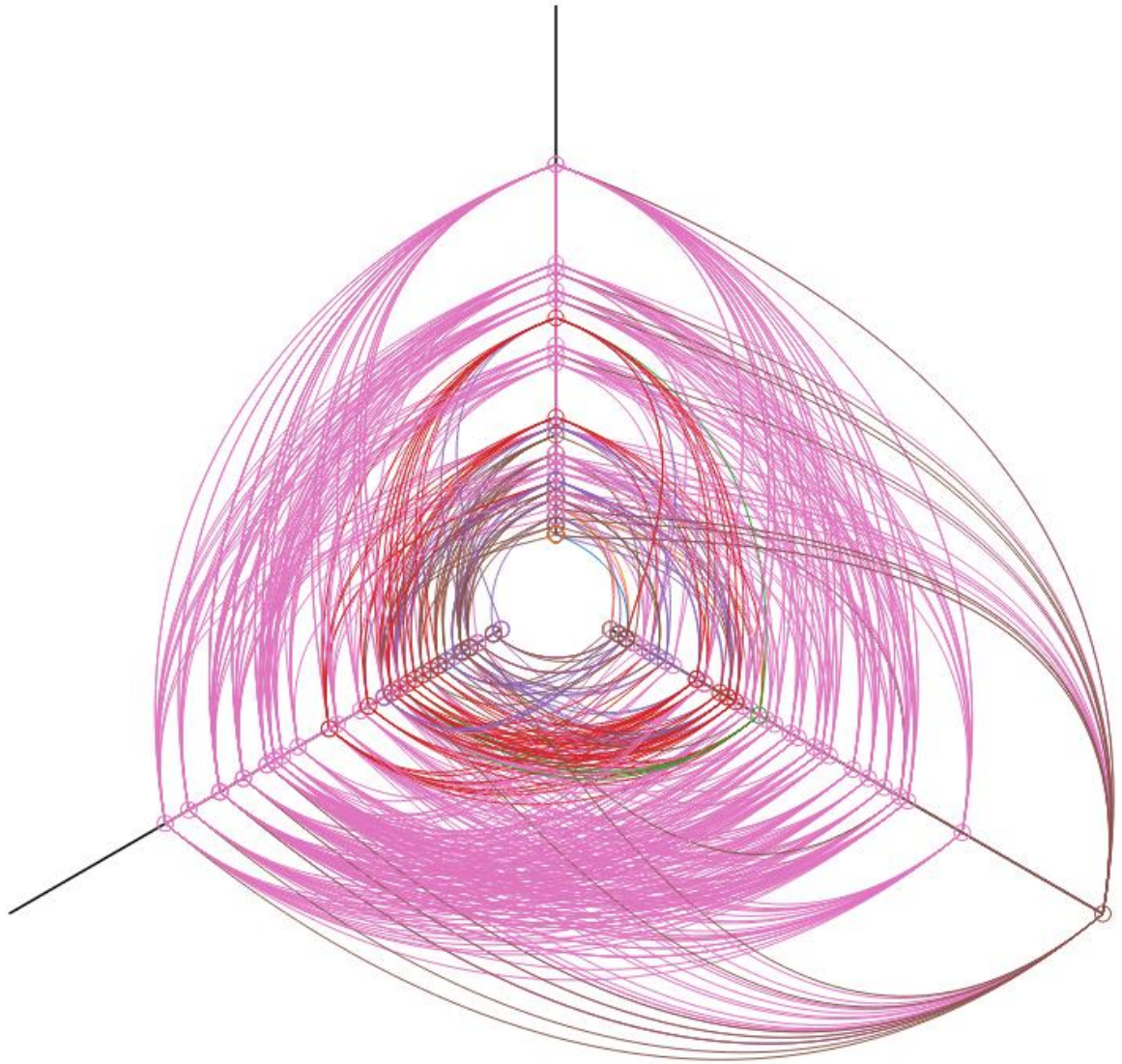
# Chapter 1

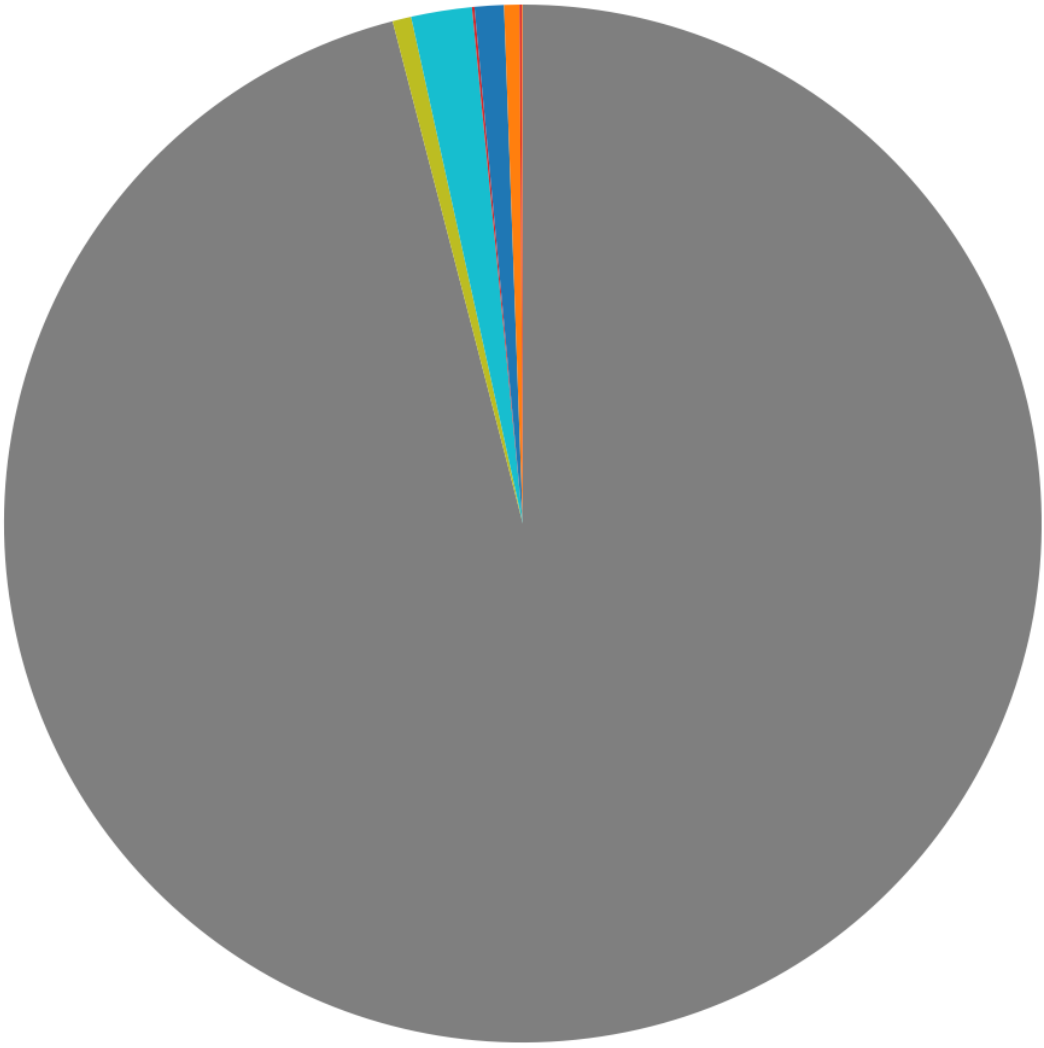
## Network Analysis – The Six Degrees of Kevin Bacon

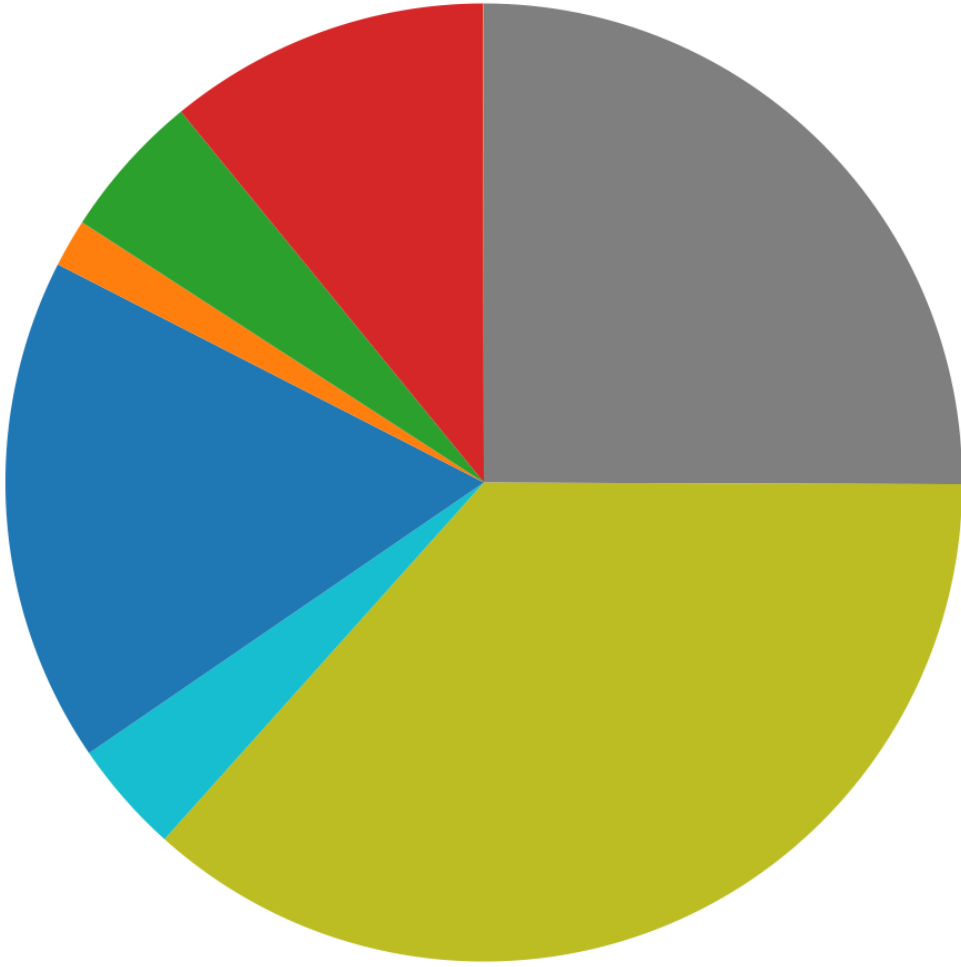






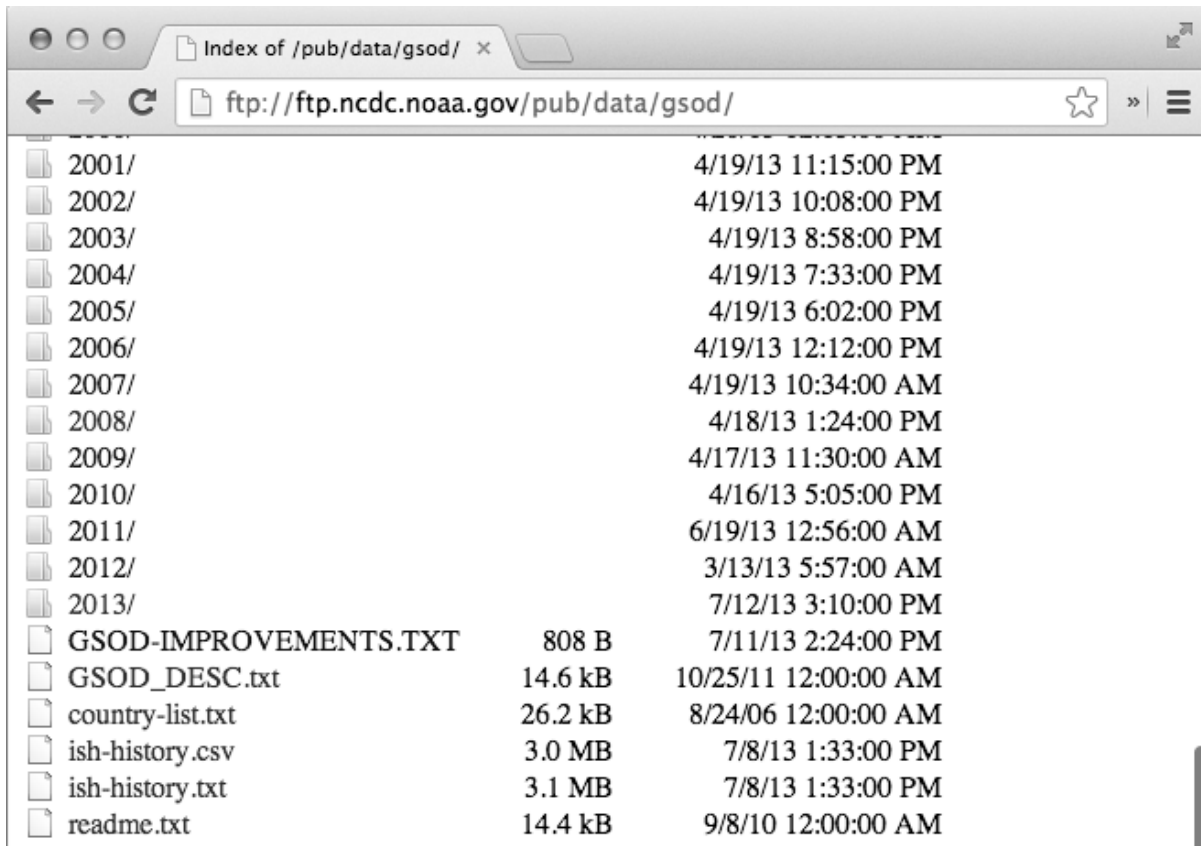






## Chapter 2

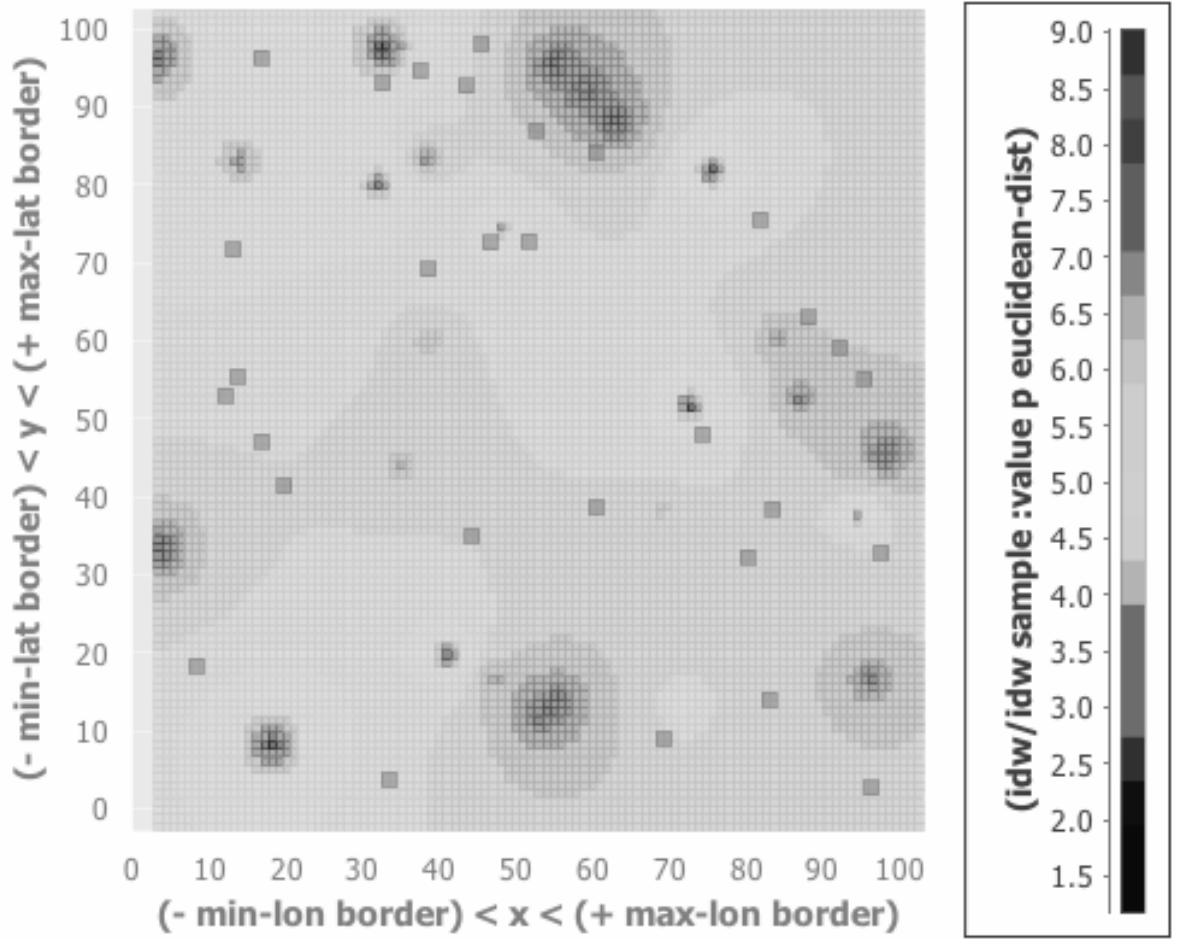
# GIS Analysis – Mapping Climate Change



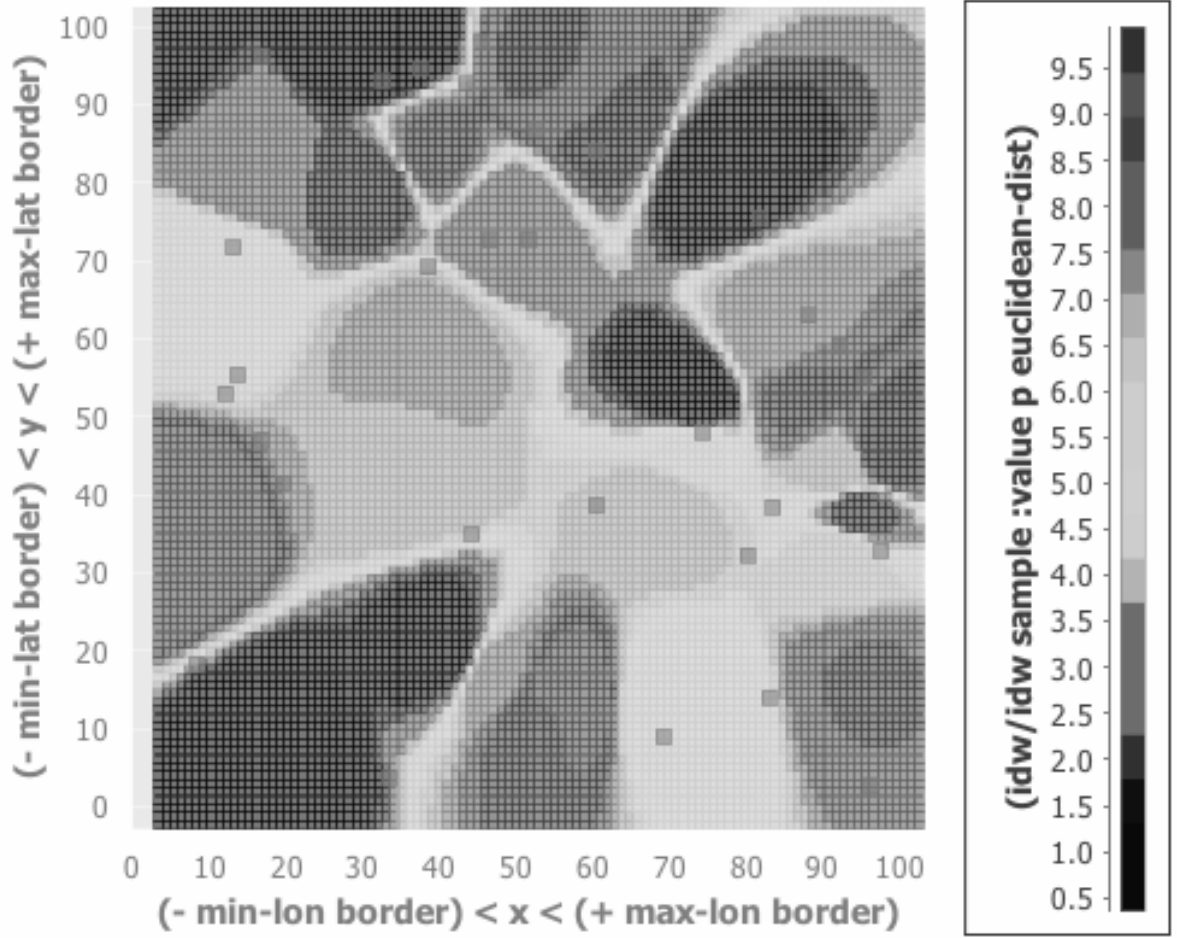
File Name	Size	Modification Date
2001/		4/19/13 11:15:00 PM
2002/		4/19/13 10:08:00 PM
2003/		4/19/13 8:58:00 PM
2004/		4/19/13 7:33:00 PM
2005/		4/19/13 6:02:00 PM
2006/		4/19/13 12:12:00 PM
2007/		4/19/13 10:34:00 AM
2008/		4/18/13 1:24:00 PM
2009/		4/17/13 11:30:00 AM
2010/		4/16/13 5:05:00 PM
2011/		6/19/13 12:56:00 AM
2012/		3/13/13 5:57:00 AM
2013/		7/12/13 3:10:00 PM
GSOD-IMPROVEMENTS.TXT	808 B	7/11/13 2:24:00 PM
GSOD_DESC.txt	14.6 kB	10/25/11 12:00:00 AM
country-list.txt	26.2 kB	8/24/06 12:00:00 AM
ish-history.csv	3.0 MB	7/8/13 1:33:00 PM
ish-history.txt	3.1 MB	7/8/13 1:33:00 PM
readme.txt	14.4 kB	9/8/10 12:00:00 AM

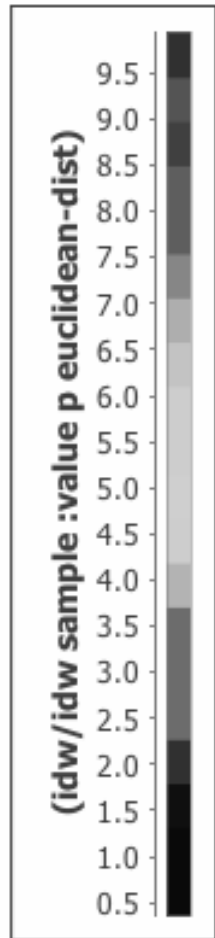
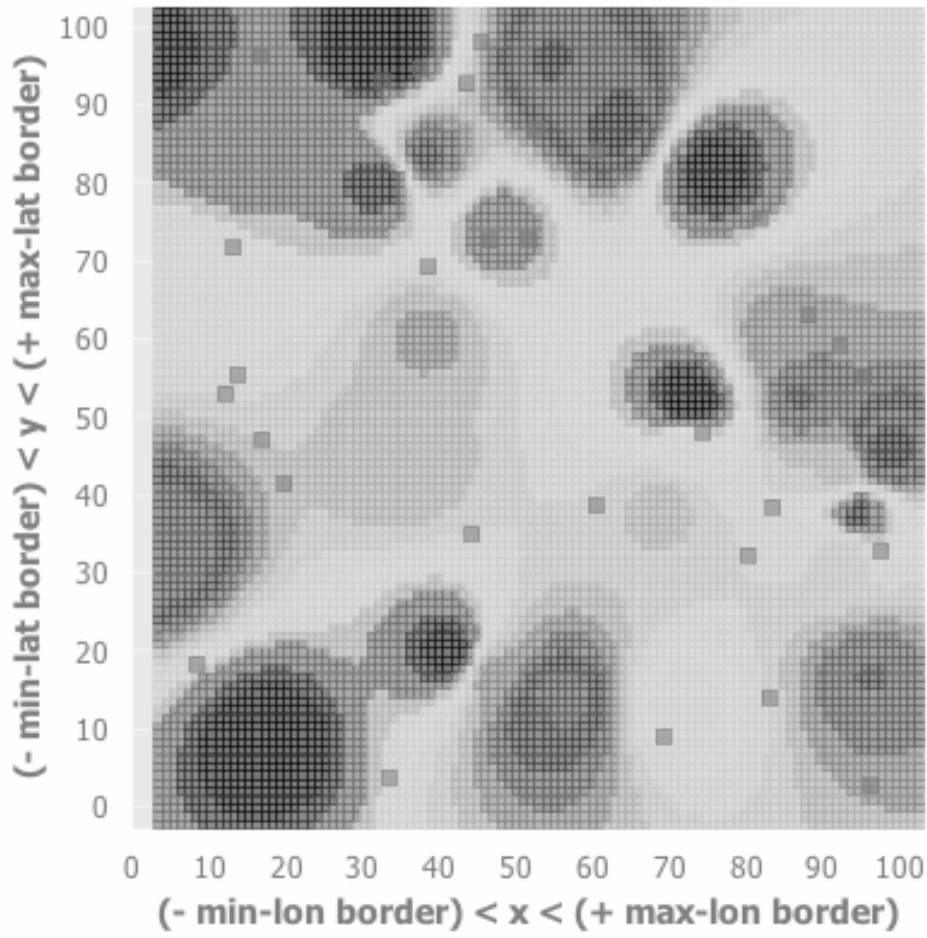
$$u(x) = \frac{\sum_{i=1}^N w_i(x) u_i}{\sum_{i=1}^N w_i(x)}$$

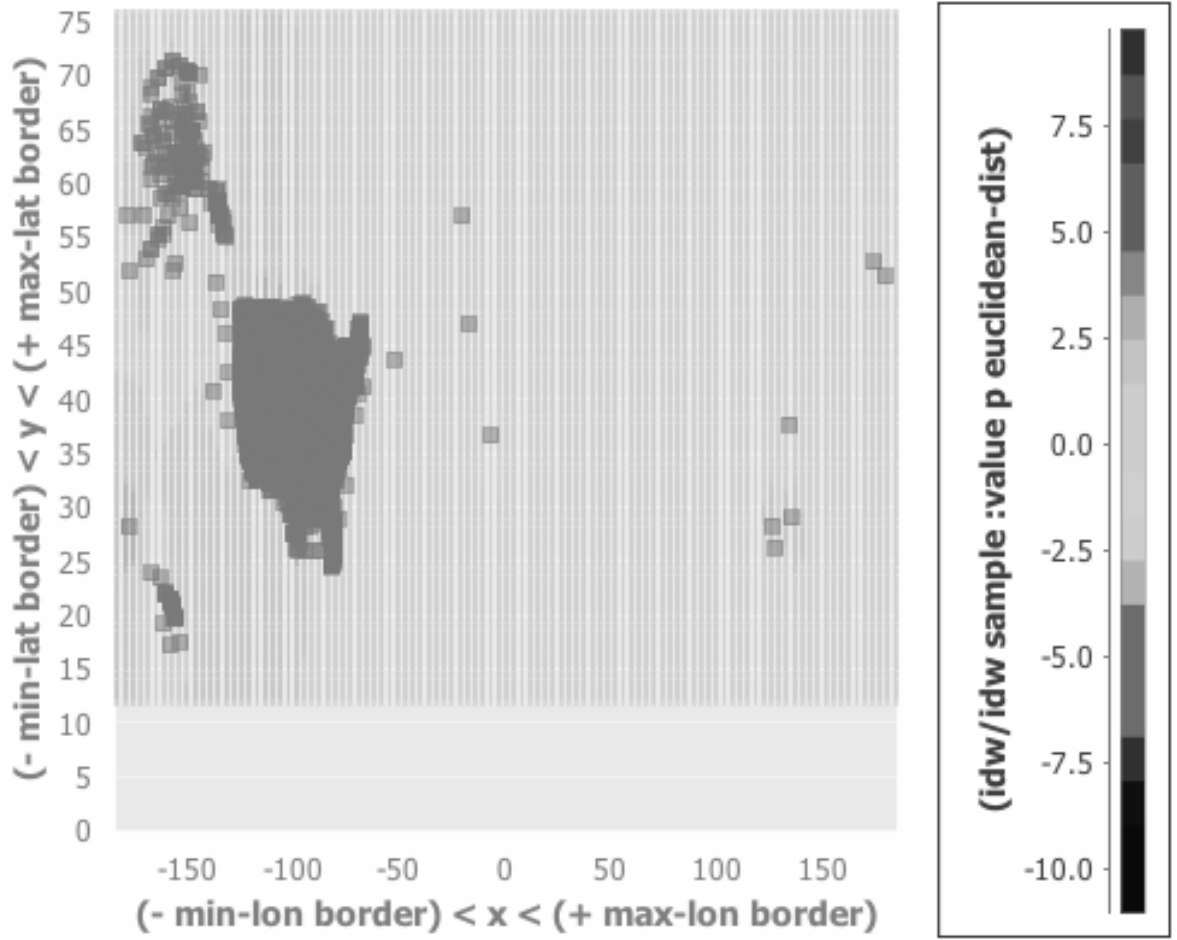
$$w_i(x) = \frac{1}{d(x, x_i)^p}$$

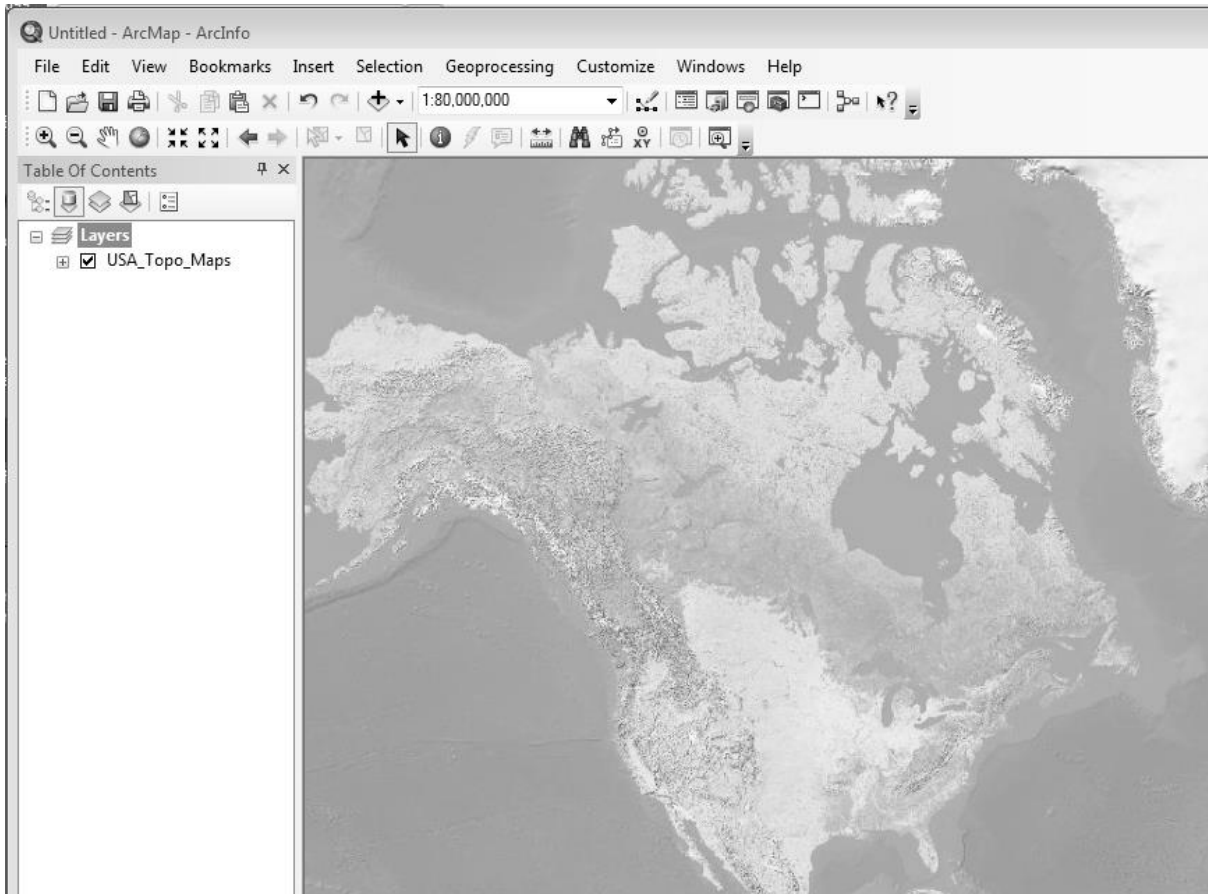












**Add XY Data** ? X

A table containing X and Y coordinate data can be added to the map as a layer

Choose a table from the map or browse for another table:

t-diffs.csv 📁

Specify the fields for the X, Y and Z coordinates:

X Field: longitude

Y Field: latitude

Z Field: z

Coordinate System of Input Coordinates

Description:

Geographic Coordinate System:  
Name: GCS\_North\_American\_1983

Show Details Edit...

Warn me if the resulting layer will have restricted functionality

OK Cancel

Layer Properties

General Source Selection Display Symbology Fields Definition Query Labels Joins & Relates Time HTML Popup

Show:

- Features
- Categories
- Quantities
  - Graduated colors
  - Graduated symbols
  - Proportional symbols
- Charts
- Multiple Attributes

**Draw quantities using color to show values.** Import...

Fields

Value: [dropdown]

Normalization: none [dropdown]

Classification

Natural Breaks (Jenks)

Classes: 5 [dropdown] Classify...

Color Ramp: [dropdown]

Symbol	Range	Label
◆	-28.712742 - -21.293633	-28.712742 - -21.293633
◇	-21.293632 - -1.092258	-21.293632 - -1.092258
◇	-1.092257 - 0.912625	-1.092257 - 0.912625
◇	0.912626 - 3.165694	0.912626 - 3.165694
◆	3.165695 - 12.954548	3.165695 - 12.954548

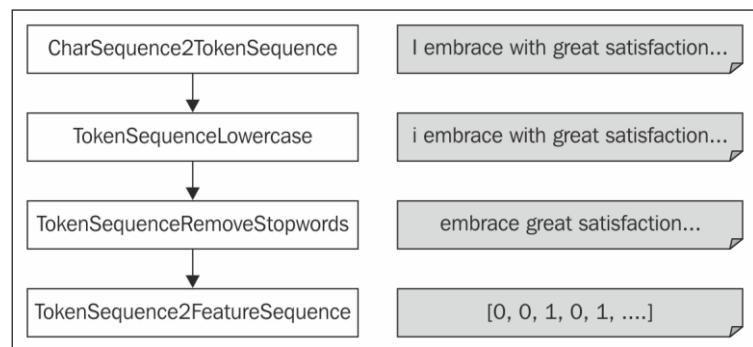
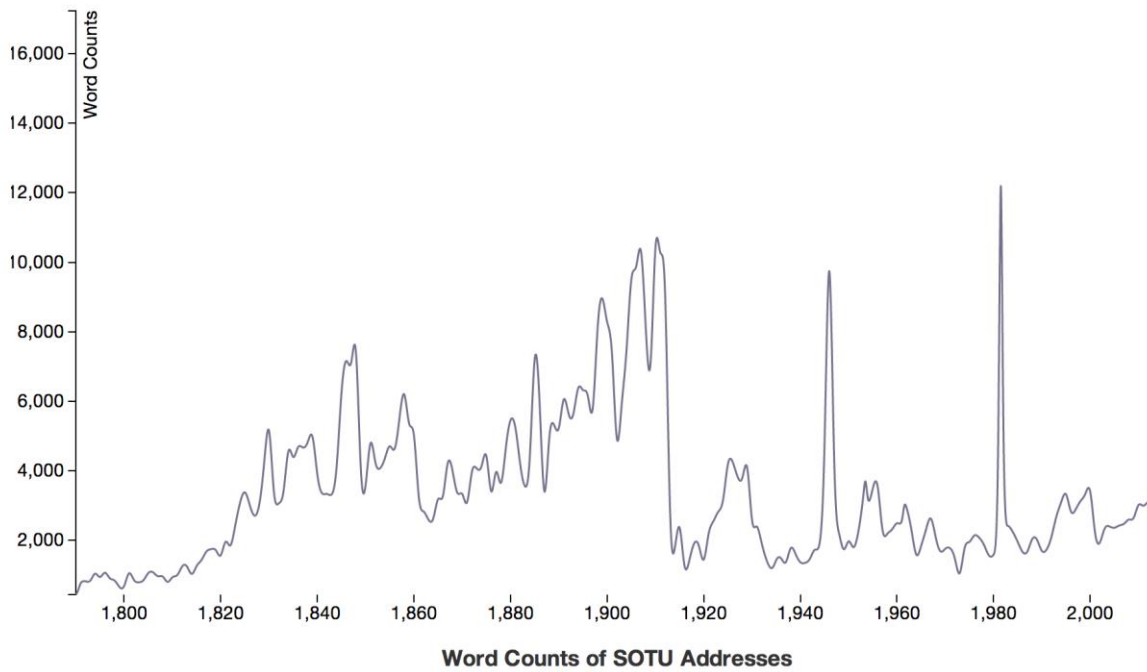
Show class ranges using feature values Advanced [dropdown]

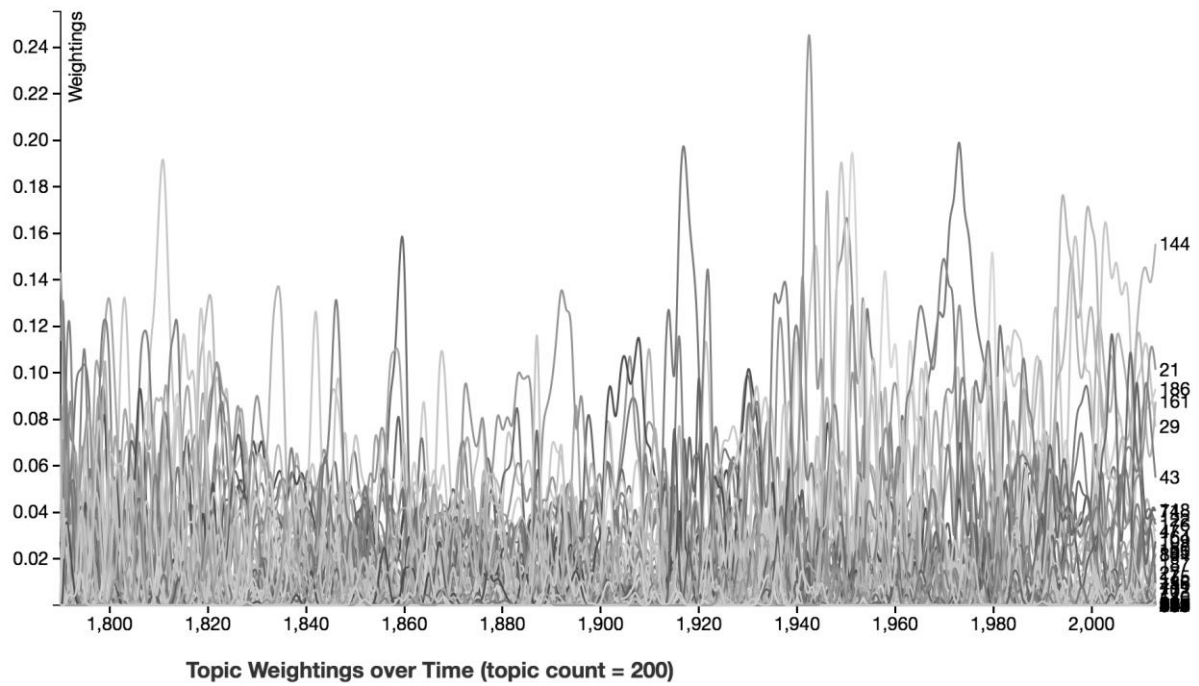
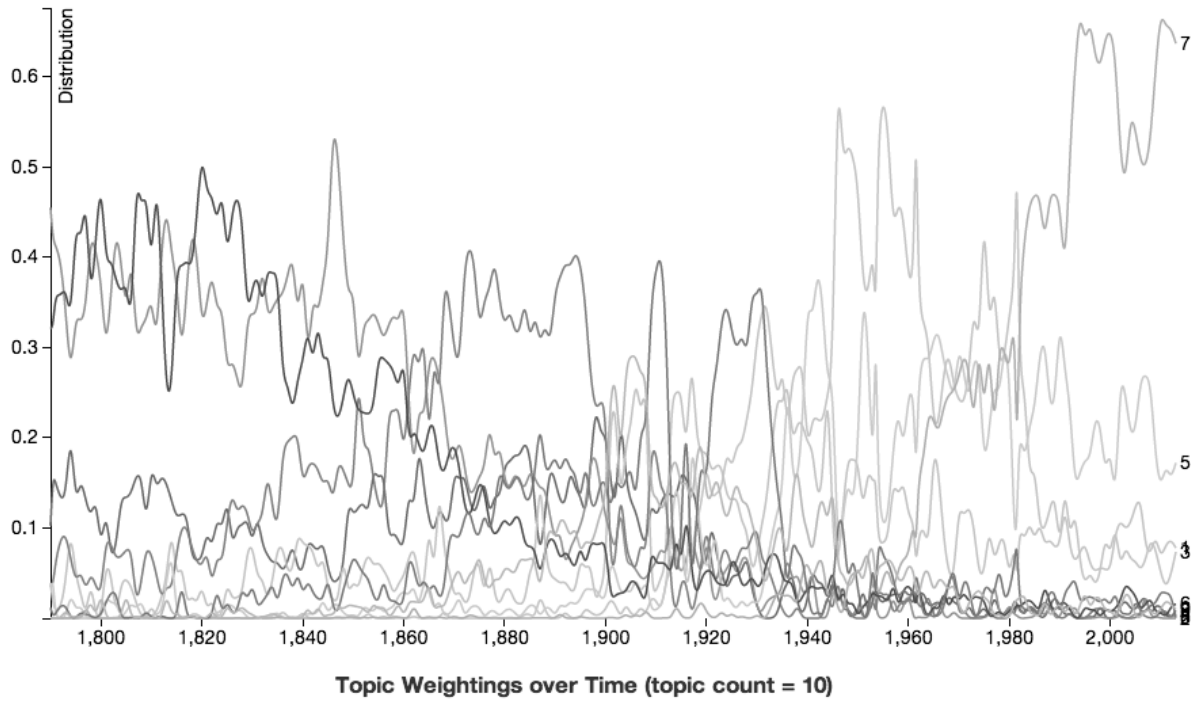
OK Cancel Apply



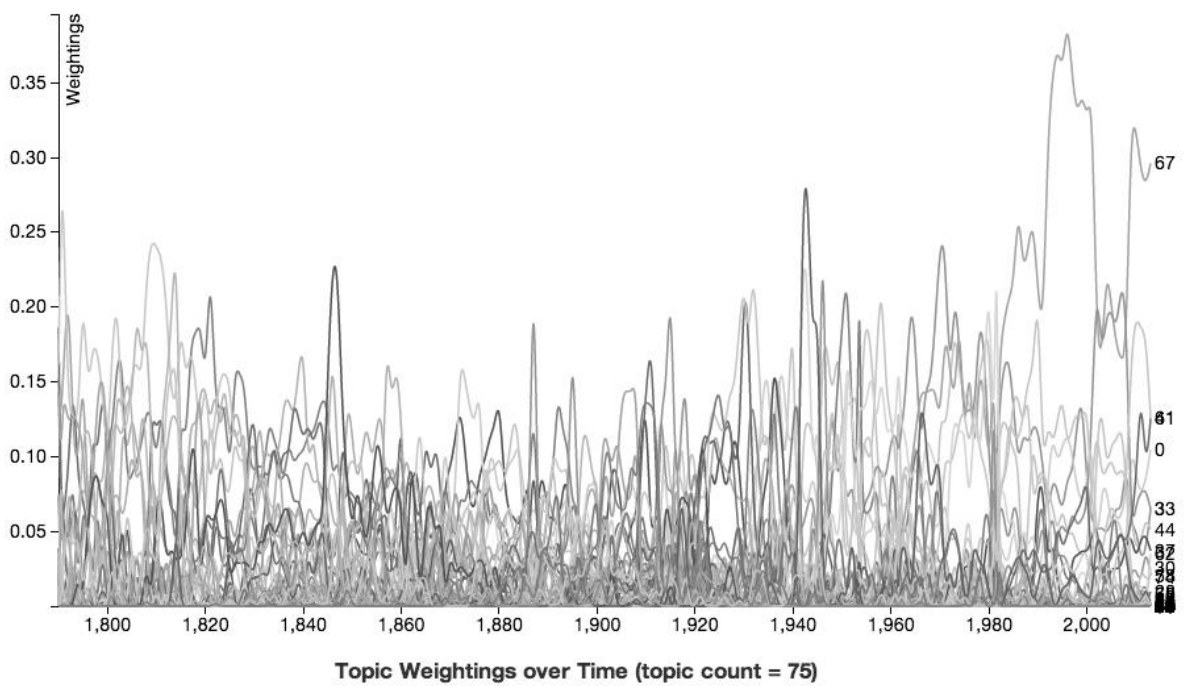
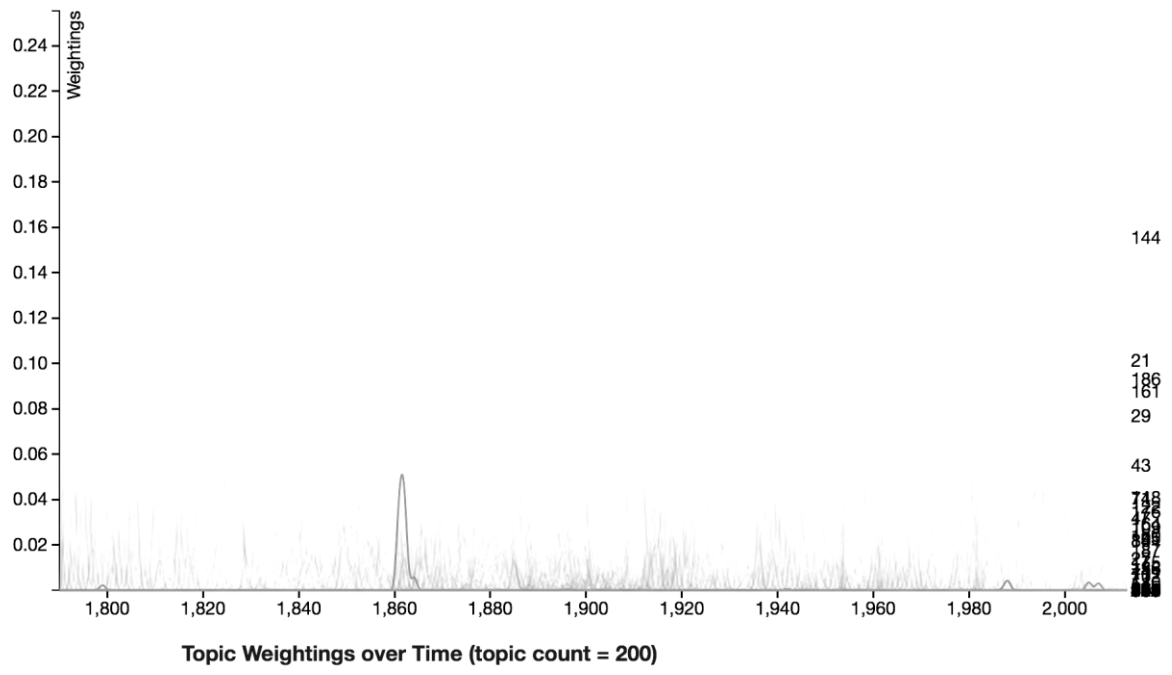
# Chapter 3

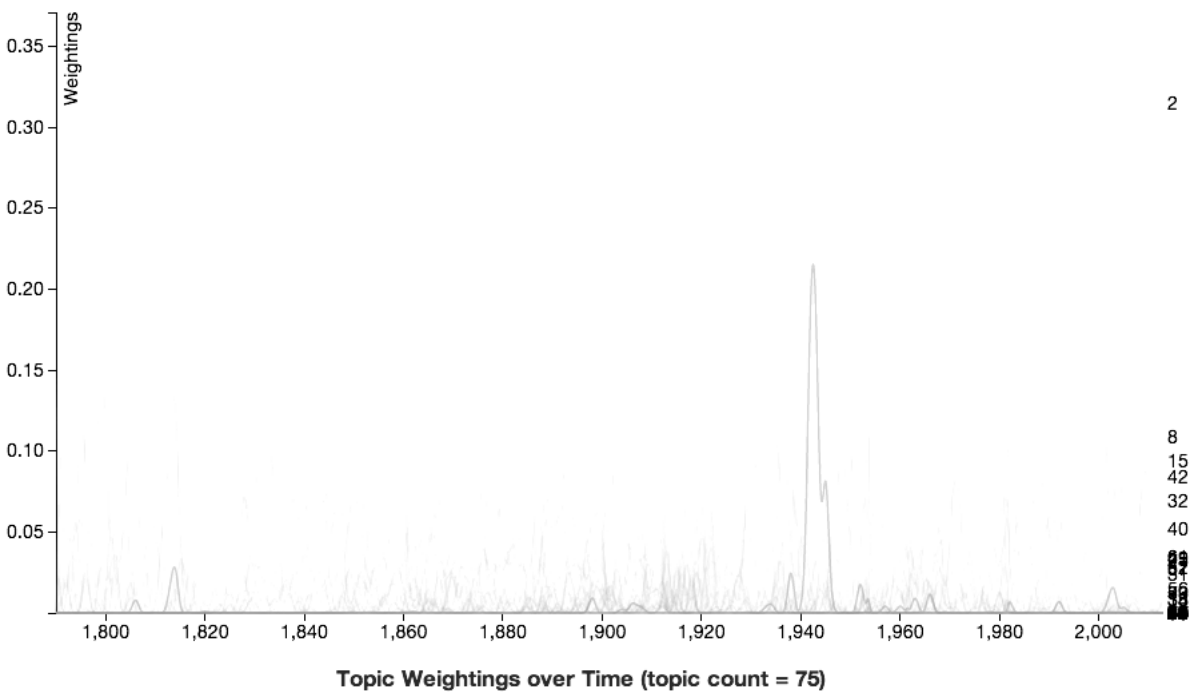
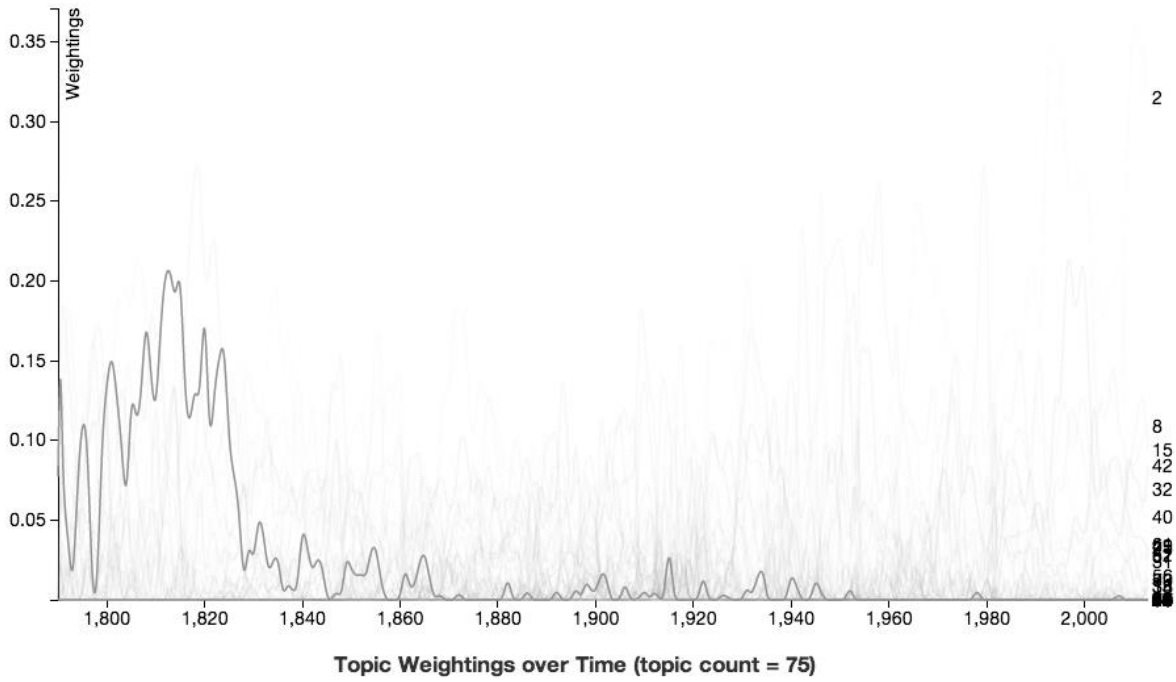
## Topic Modeling – Changing Concerns in the State

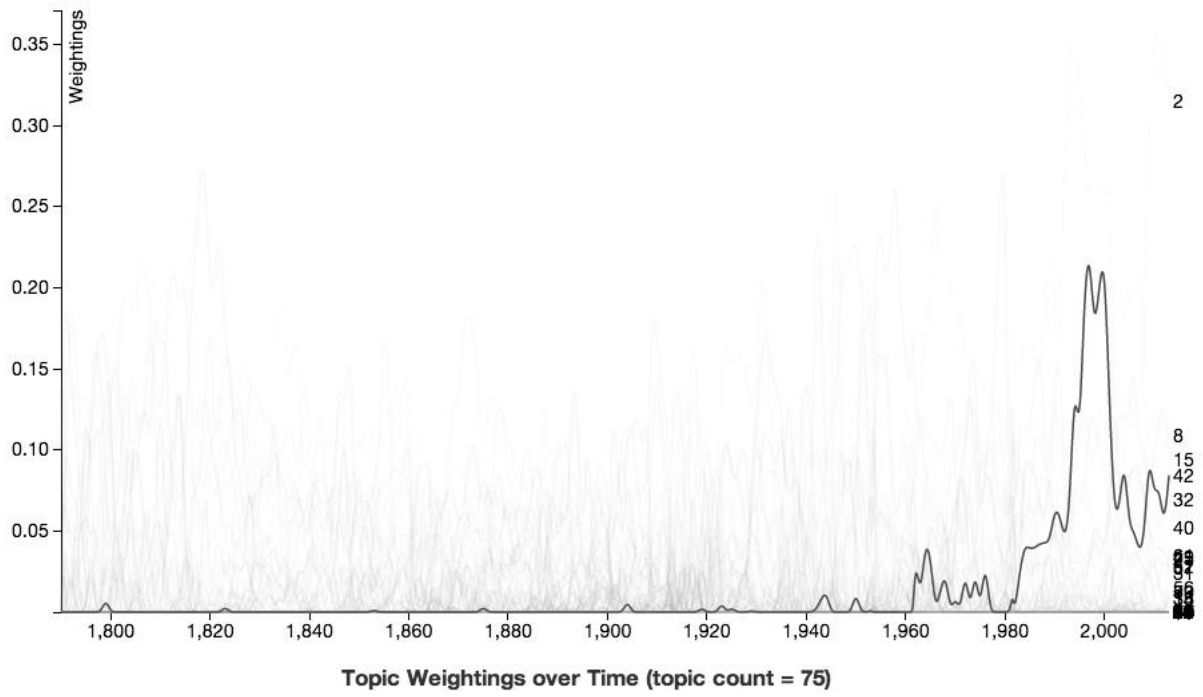












# Chapter 4

## Classifying UFO Sightings

infochimps

Solutions | Infochimps Cloud | Resources | Community | Company

HOME / DATA MARKETPLACE / DATASETS / 60,000+ DOCUMENTED UFO SIGHTINGS WITH TEXT DESCRIPTIONS AND METADATA

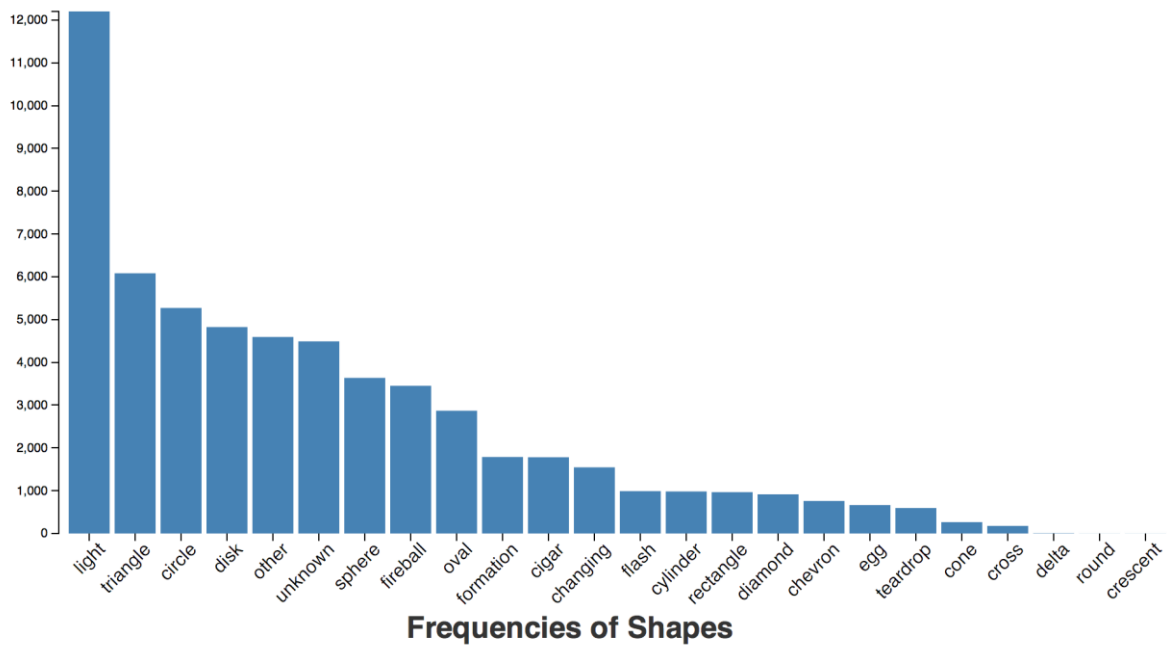
### Dataset

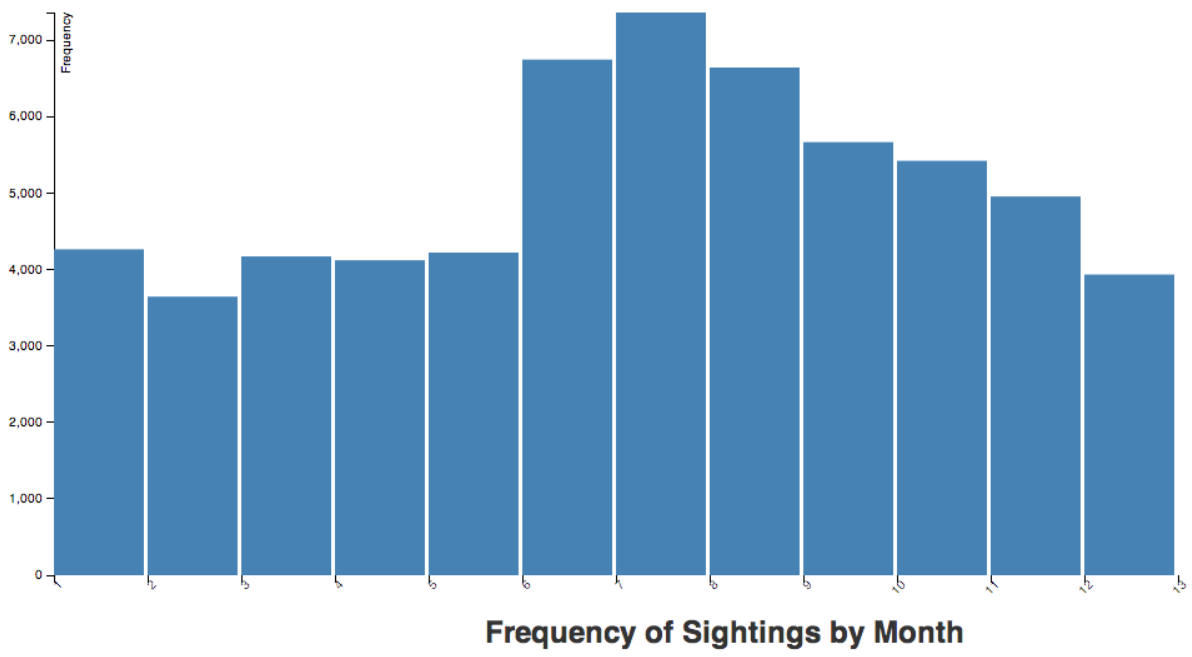
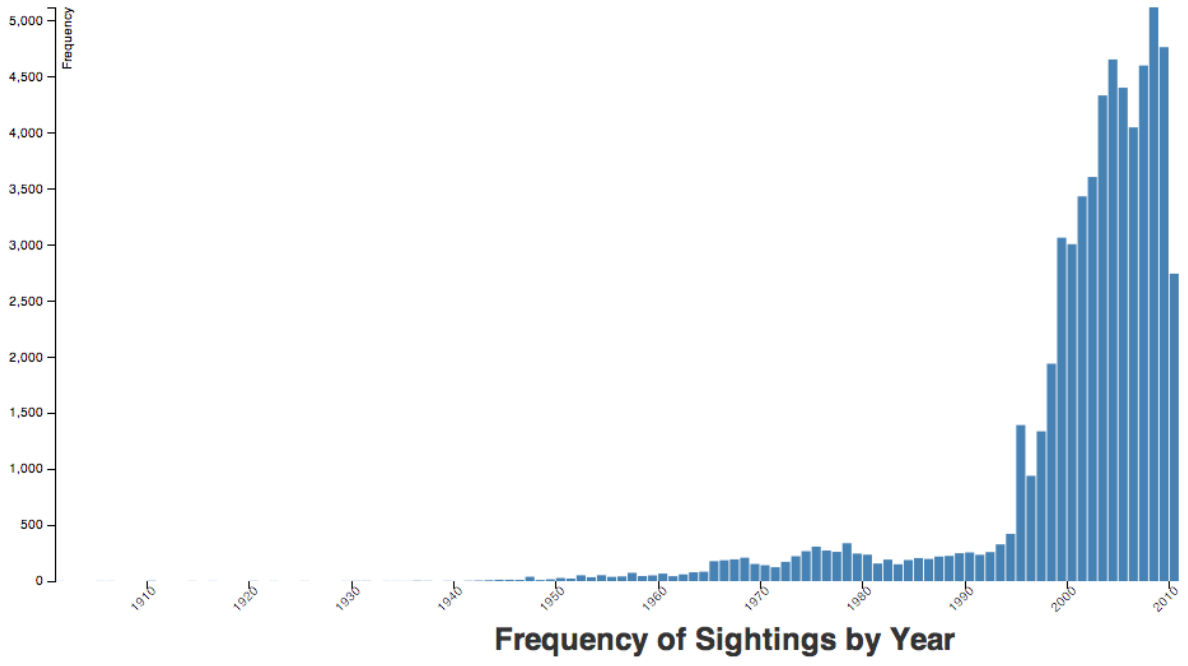
60,000+ DOCUMENTED UFO SIGHTINGS WITH TEXT DESCRIPTIONS AND METADATA  
Added By **Infochimps**

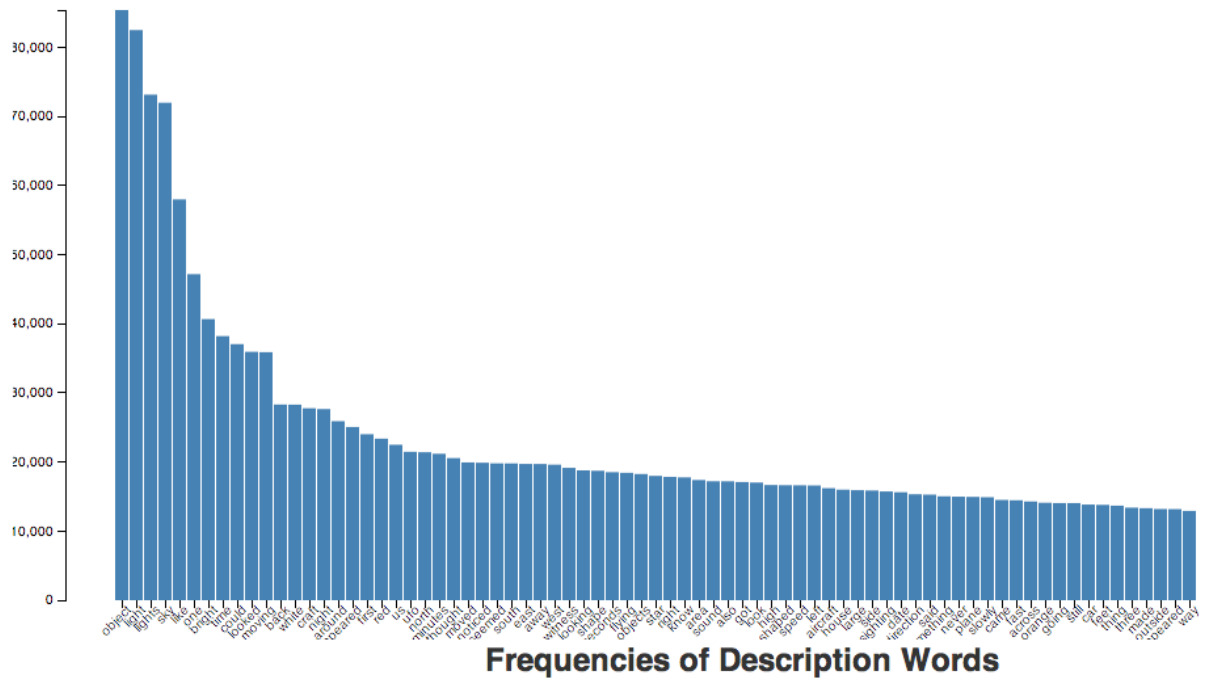
**The service for this API has ceased**  
Our apologies for the inconvenience this may cause.  
You can find a download of the data set for this API on this page  
Search UFO data across more than 60,000 accounts of UFO sightings, including

**DOWNLOAD** 78.1 MB  
BZ2 (ZIP Compressed) [Download >](#)

REQUEST A DEMO »

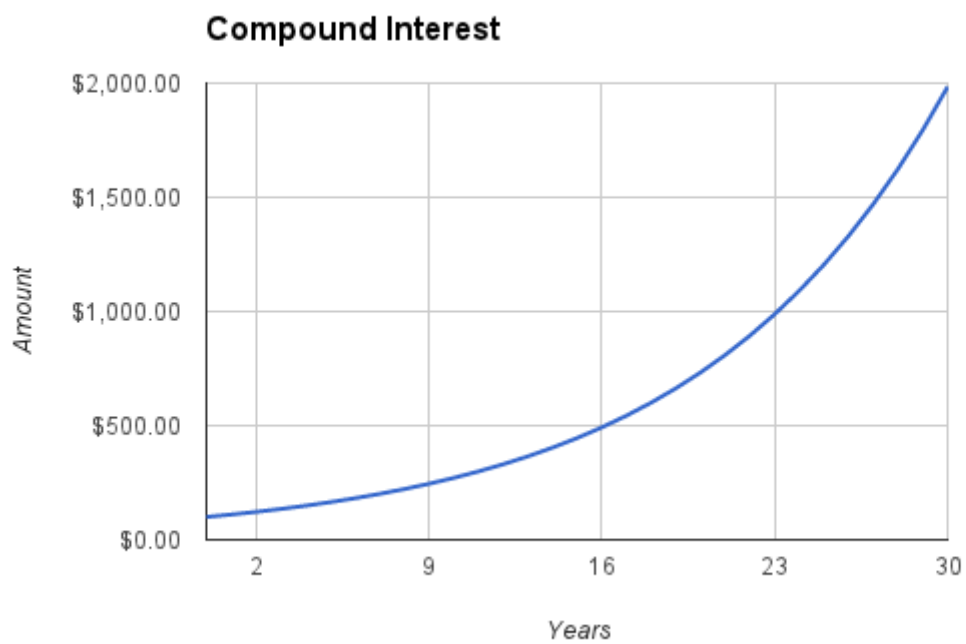
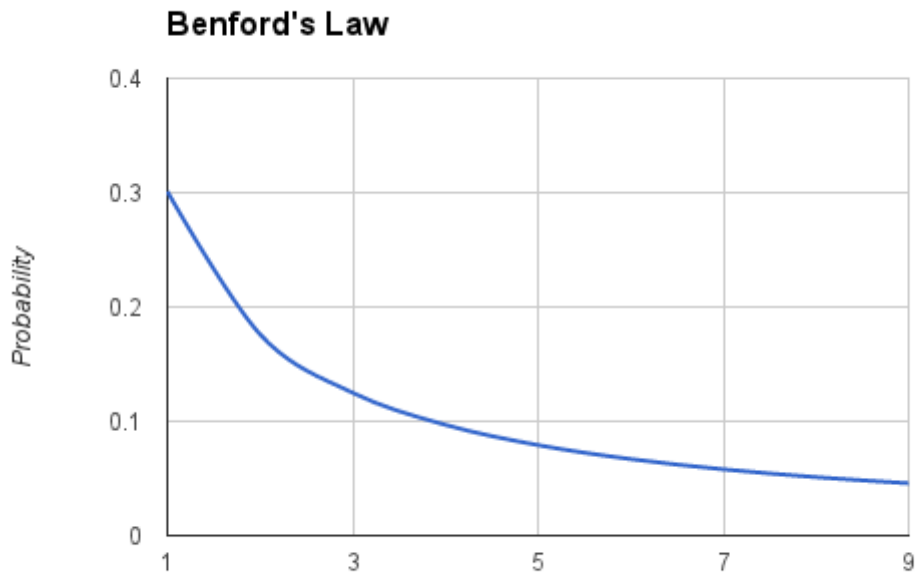




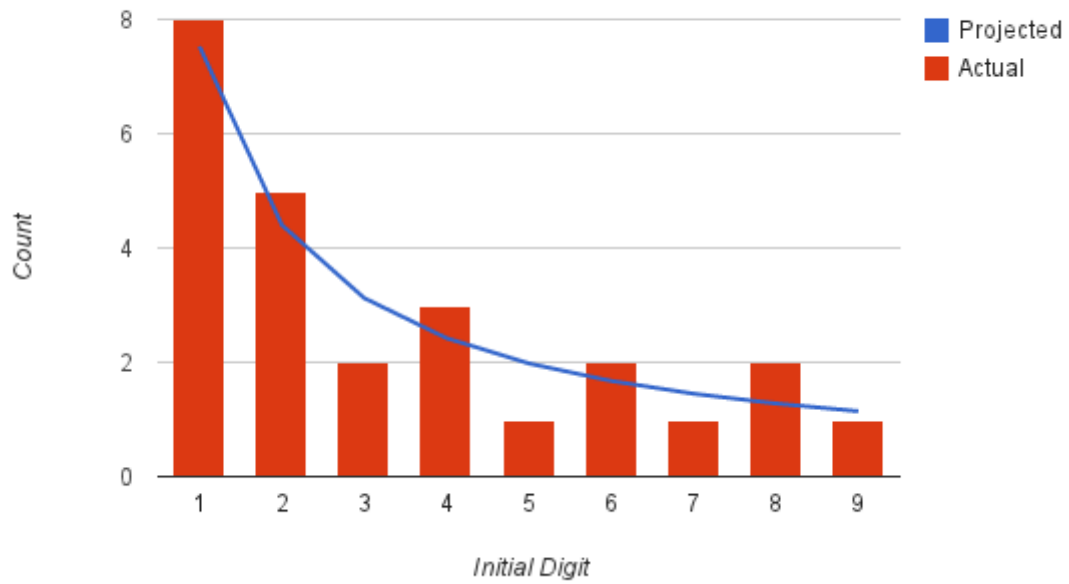


# Chapter 5

## Benford's Law – Detecting Natural



## Benford's Law: Compounding Interest



$$X^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}$$

After selecting variables, please click one of the above options

**DIMENSION FILTERS**

Hierarchy  
Collapse all | Expand all

Year

**COUNTRY** (Available:214 | Selected:214)

**SERIES** (Available:60 | Selected:1)

**TIME** (Available:53 | Selected:53)

SELECT YEARS | FUNCTIONS

Availability Range: Year 1960 - 2012

Please drag the slider handle below to select the range

Select all | Unselect all | Invert selection

2012

**YOUR CURRENT SELECTION**

▼ DATABASE  
World Development Indicators  
Change database

► COUNTRY (214)

► SERIES (1)

▼ TIME (53)  
Remove all | View all  
Only 20 items listed below

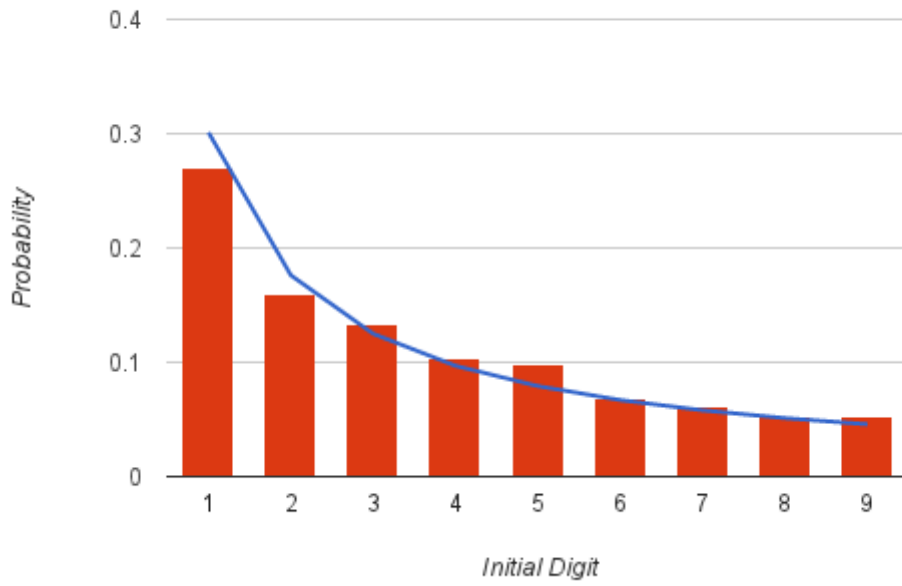
1960

1961

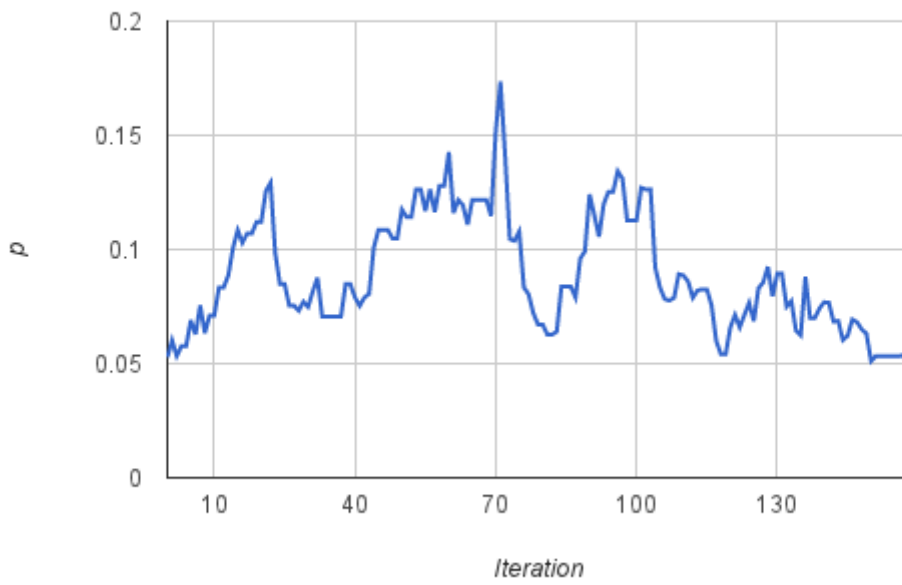
1962



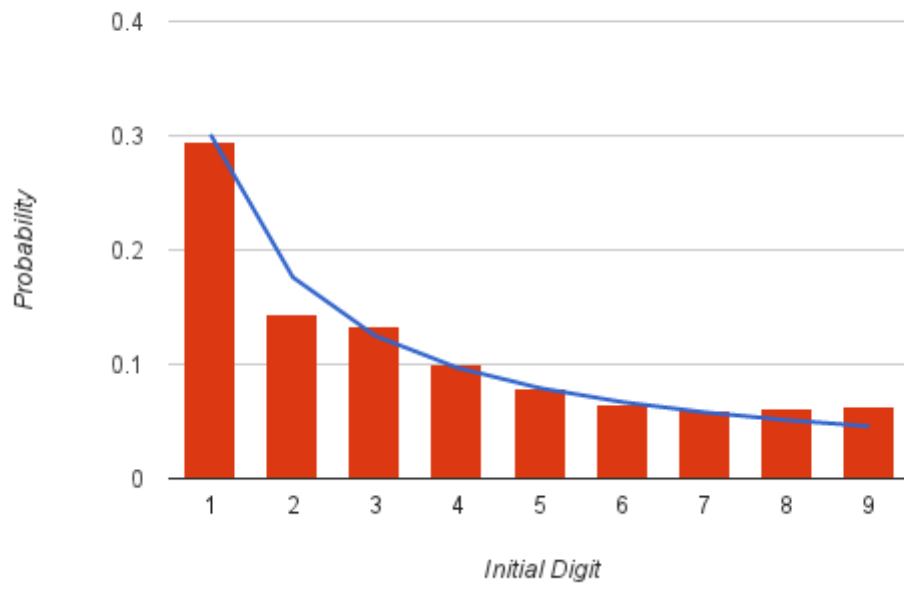
### Benford's Law: World Population



### Becoming Fraudulent

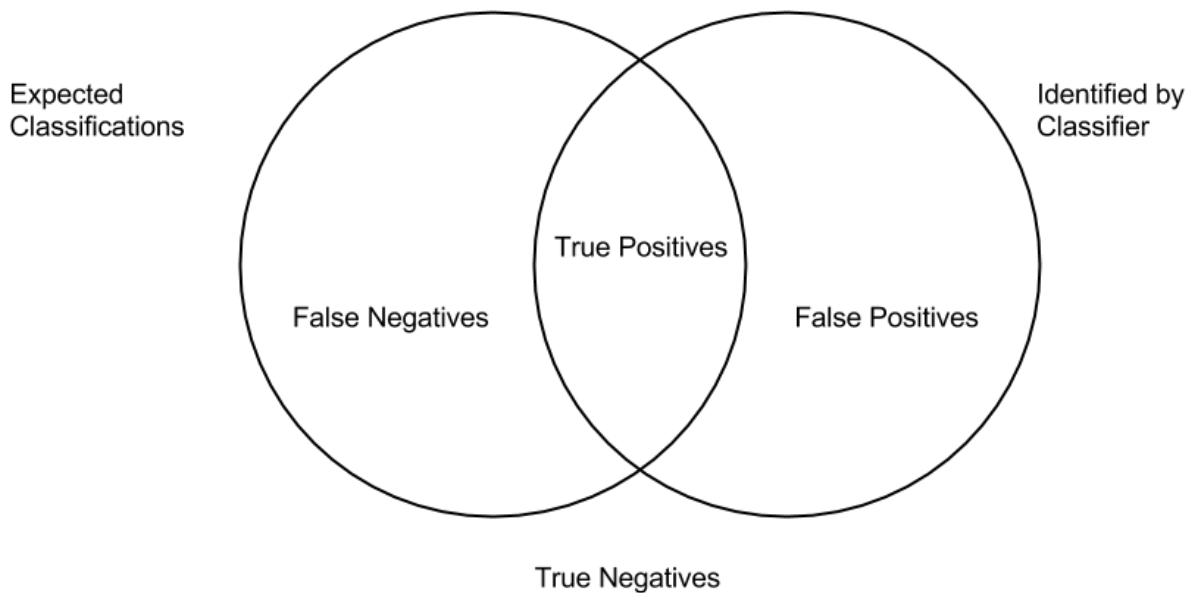


## Benford's Law: Random Swaps



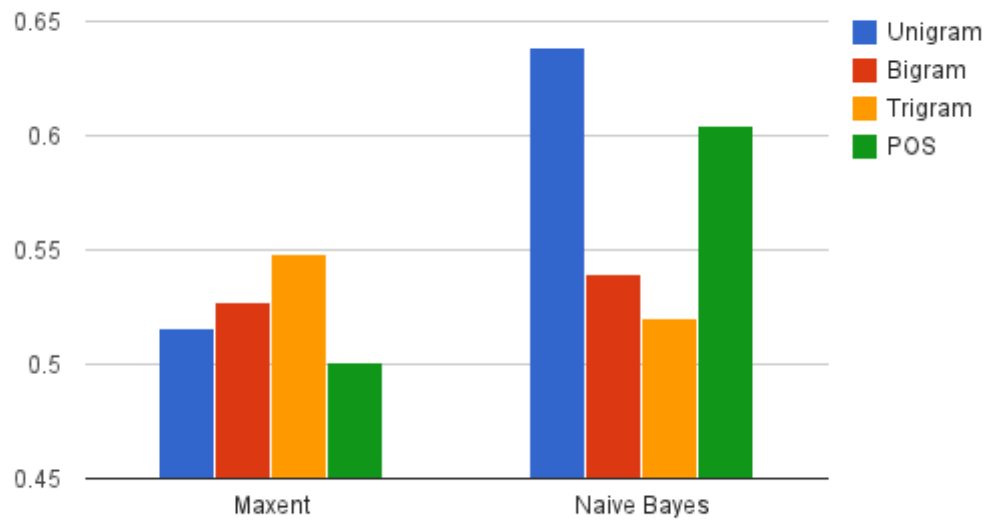
## Chapter 6

# Sentiment Analysis – Categorizing Hotel Reviews

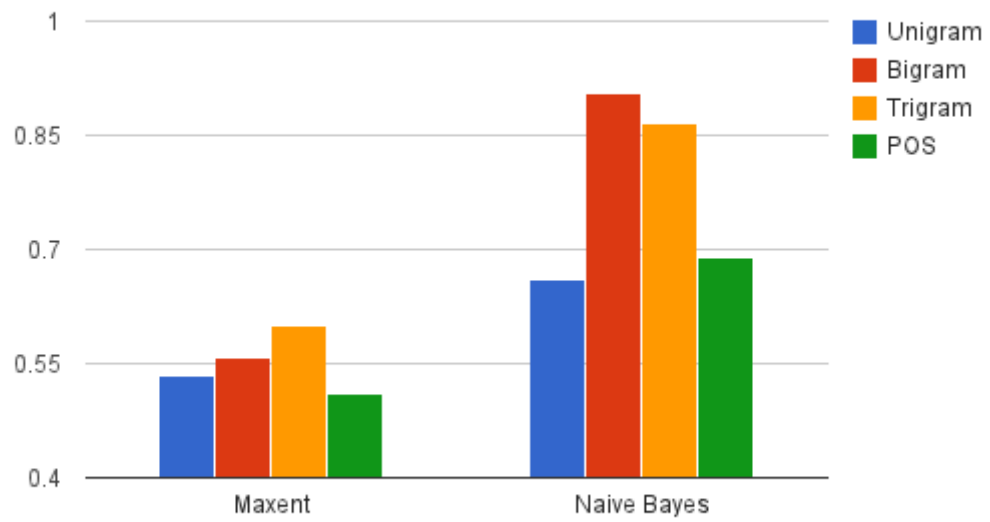


$$p(C|F) = \frac{p(C)p(F|C)}{p(F)}$$

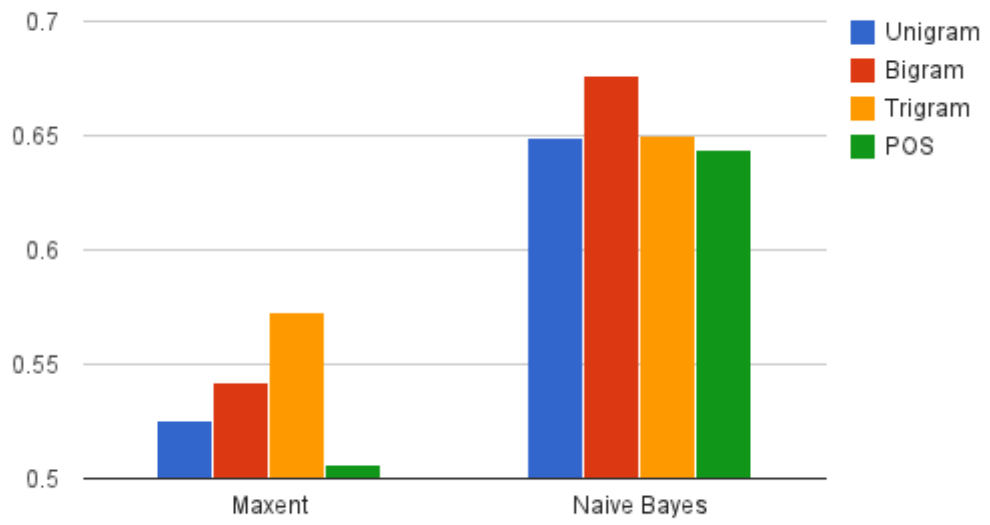
### Precision



### Recall

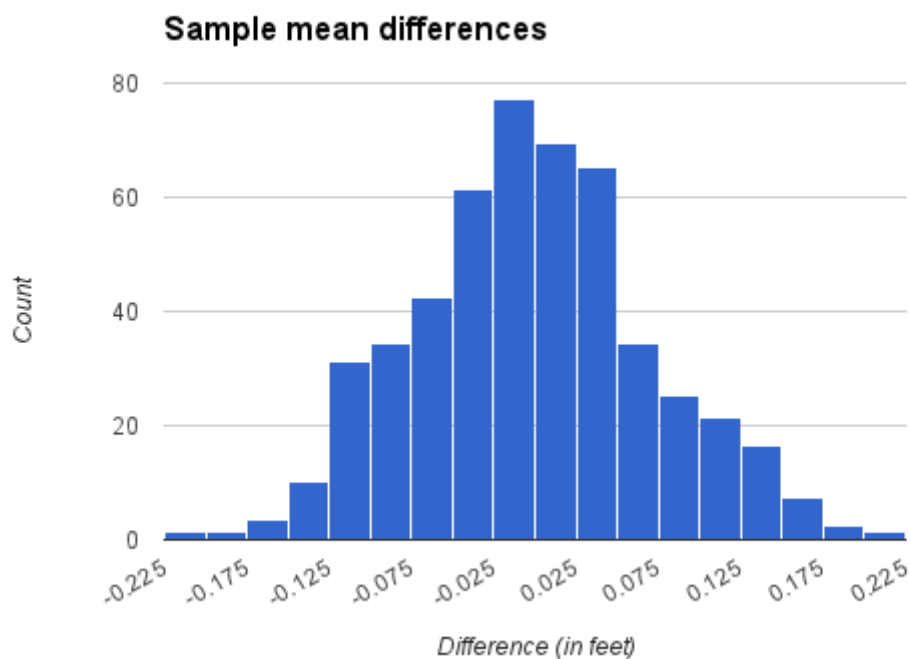
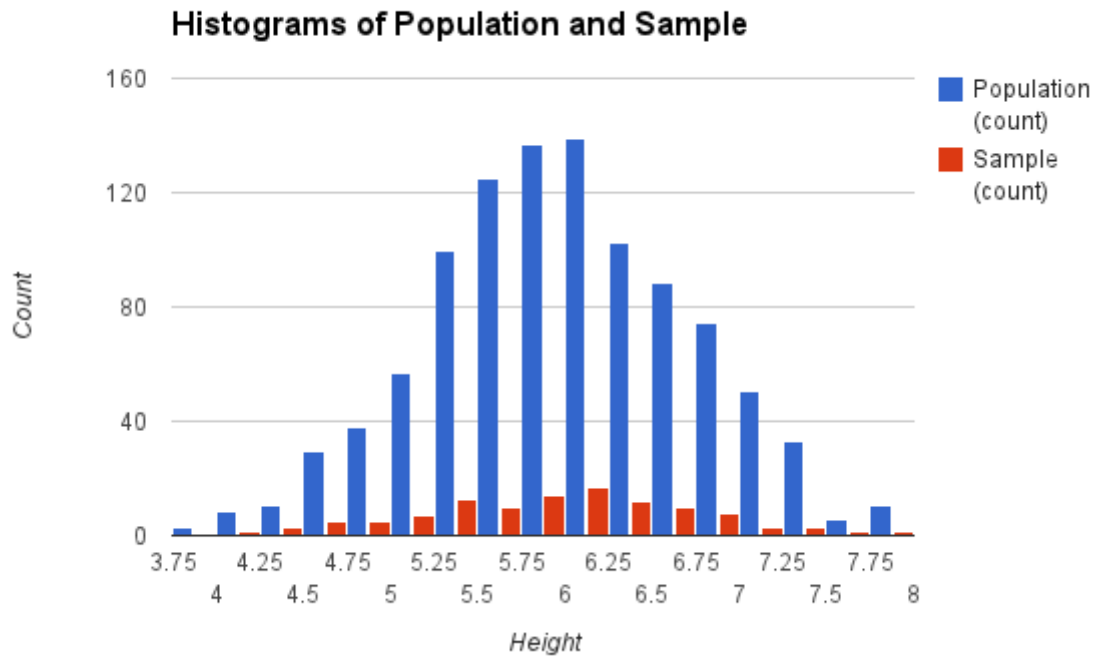


### F-measure



# Chapter 7

## Null Hypothesis Tests – Analyzing Crime Data



$$H_0 : \mu \geq \mu_0$$

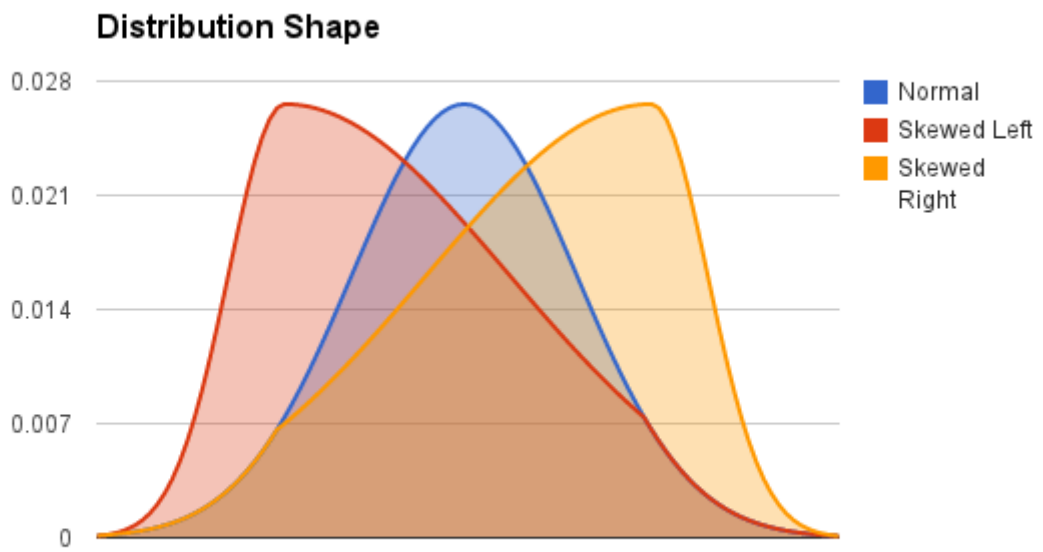
$$H_a : \mu < \mu_0$$

$$H_0 : \mu \leq \mu_0$$

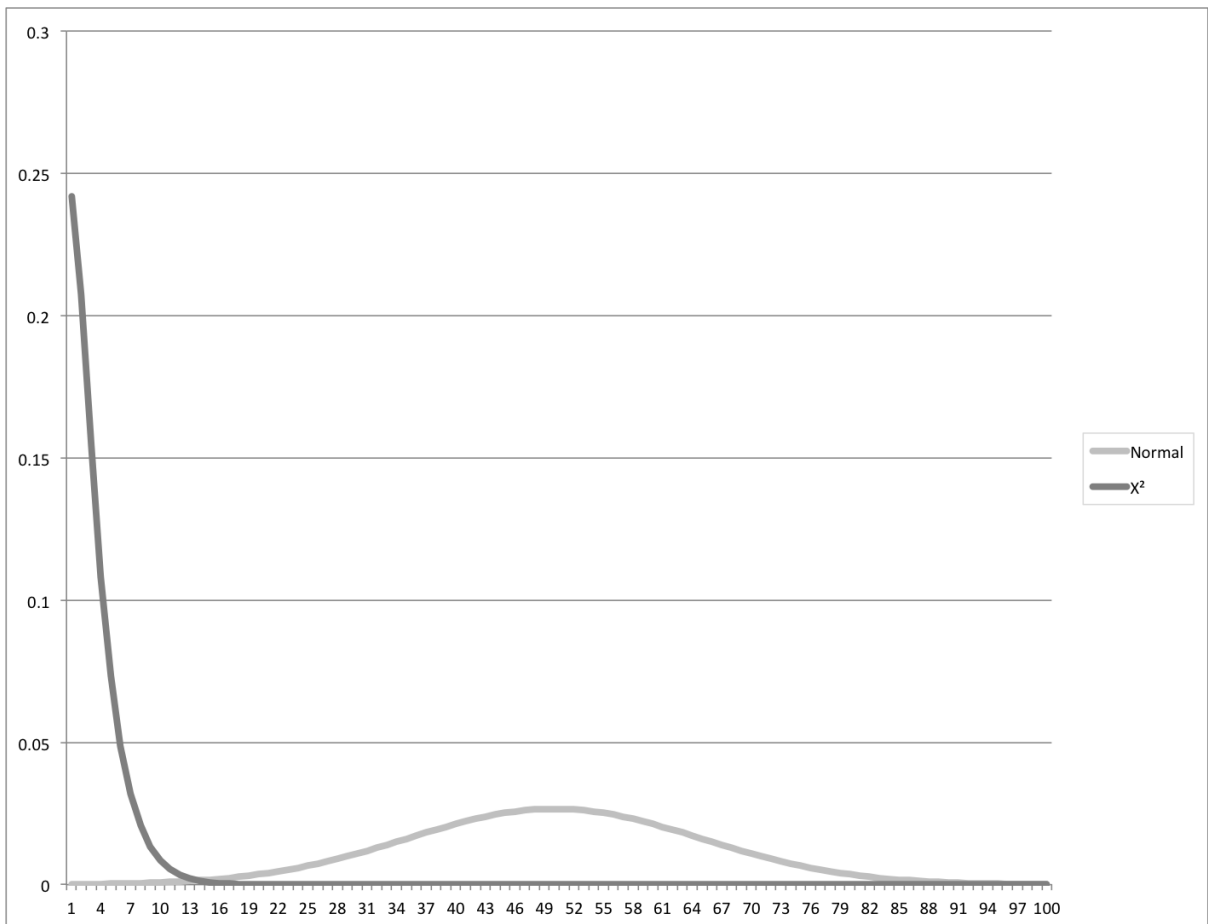
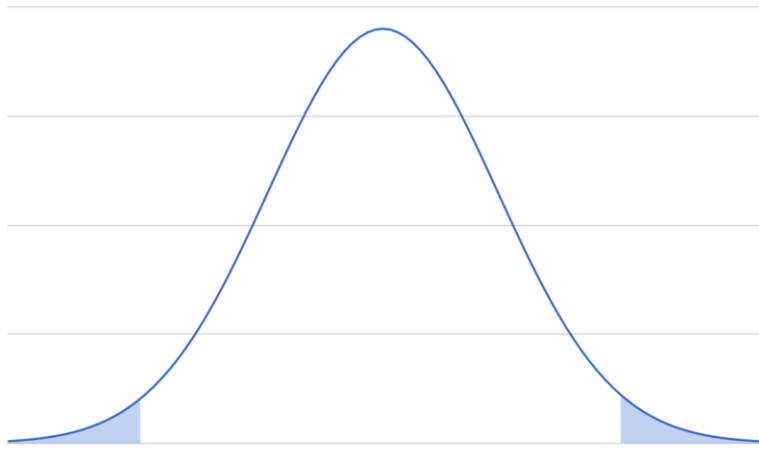
$$H_a : \mu > \mu_0$$

$$H_0 : \mu = \mu_0$$

$$H_a : \mu \neq \mu_0$$



Critical Regions





# Data

[By Country](#)[By Topic](#)[Indicators](#)[Data Catalog](#)[Microdata](#)[Blog](#)

This page in [English](#) | [Español](#) | [Français](#) | [العربية](#) | [中文](#)

## Indicators

### Find an indicator

Agricultural land (% of land area)

Arable land (% of land area)

Forest area (% of land area)

Land area where elevation is below 5 meters (% of total land area)

Land area (sq. km)

Permanent cropland (% of land area)

Population density (people per sq. km of land area)

Terrestrial protected areas (% of total land area)

Road density (km of road per 100 sq. km of land area)

Agricultural irrigated land (% of total agricultural land)

## Land area (sq. km)

[DATABANK](#)[DOWNLOAD DATA](#)[SHARE](#)[EXCEL](#)[XML](#)[CSV](#)

Land area is a country's total area, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones. In most cases the definition of inland water bodies includes major rivers and lakes.

Food and Agriculture Organization, electronic files and web site.

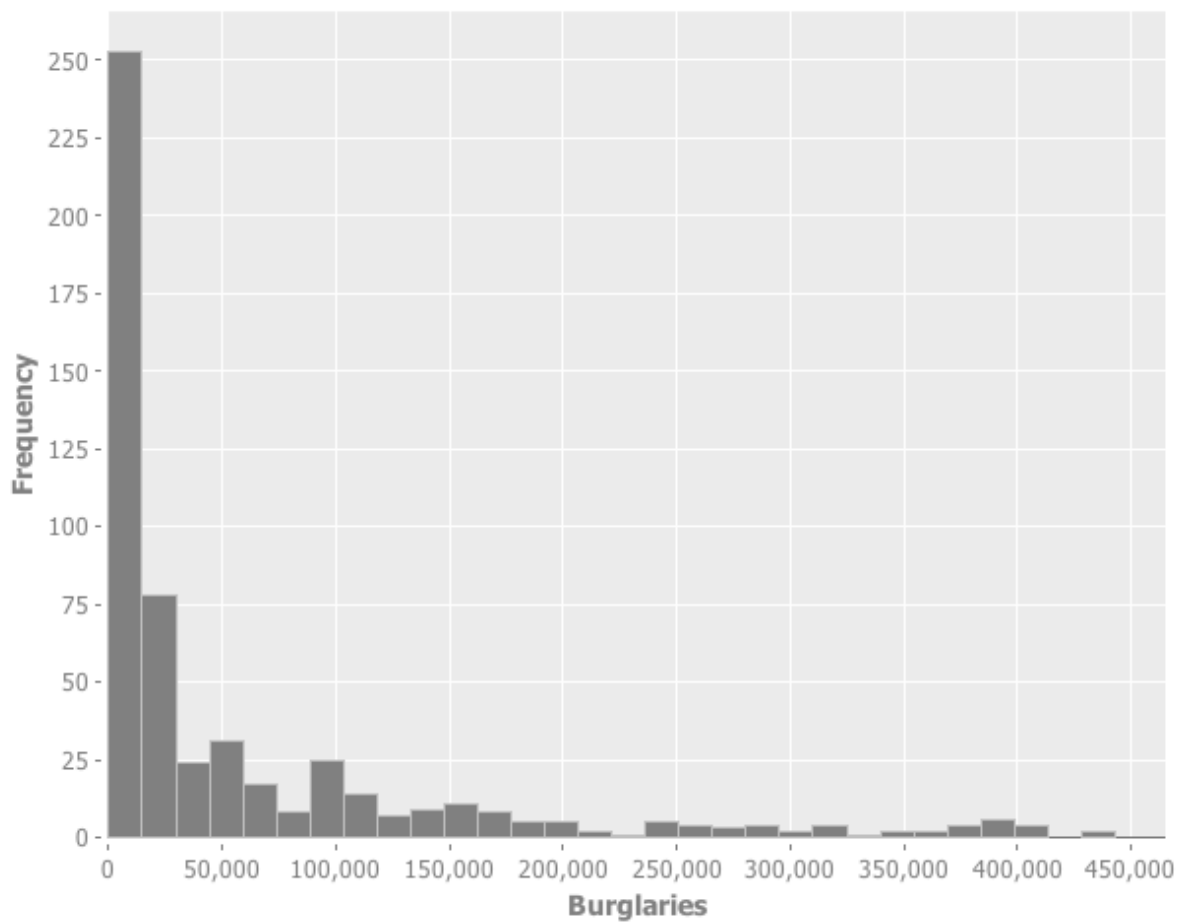
### Name

▼	ag.lnd
	[Content_Types].xml
	ag.lnd.totl.k2_Indicator_en_csv_v2.csv
	ag.lnd.totl.k2_In...ntry_en_csv_v2.csv
	ag.lnd.totl.k2_In...ator_en_csv_v2.csv
▼	ny.gnp
	[Content_Types].xml
	ny.gnp.pcap.cd_l...tor_en_csv_v2.csv
	ny.gnp.pcap.cd_l...try_en_csv_v2.csv
	ny.gnp.pcap.cd_l...tor_en_csv_v2.csv
▼	unodc-data
	CTS_Assault.xls
	CTS_Burglary.xls
	CTS_Kidnapping.xls
	CTS_Motor_vehicle_theft.xls
	CTS_Robbery.xls
	CTS_Sexual_violence.xls
	CTS_Theft.xls
	Drug_related_crimes.xls

	A	D	E	F
10				
11				
12				
13				
14	<b>Region</b>	<b>Sub-region II</b>	<b>Country/territory</b>	<b>2003</b>
15	Africa	Eastern Africa	Burundi	
16	Africa		Kenya	
17	Africa		Mauritius	105
18	Africa		Mozambique	
19	Africa		Uganda	19,491
20	Africa		Zimbabwe	78,464
21	Africa	Middle Africa	Cameroon	2,998
22	Africa		Sao Tome and Principe	1
23	Africa	Northern Africa	Algeria	23,050
24	Africa		Egypt	
25	Africa		Morocco	48,662
26	Africa	Southern Africa	Botswana	
27	Africa		Lesotho	

Incanter Plot

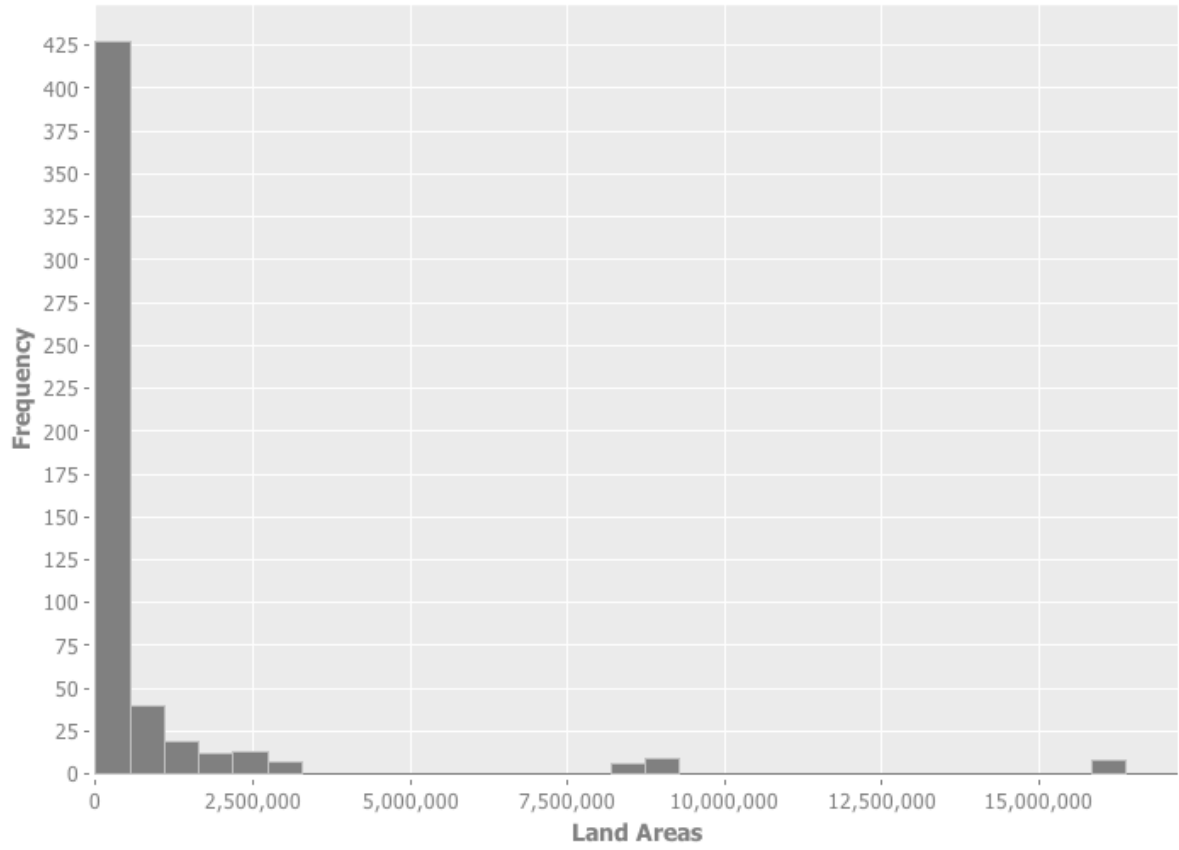
### Burglary Counts



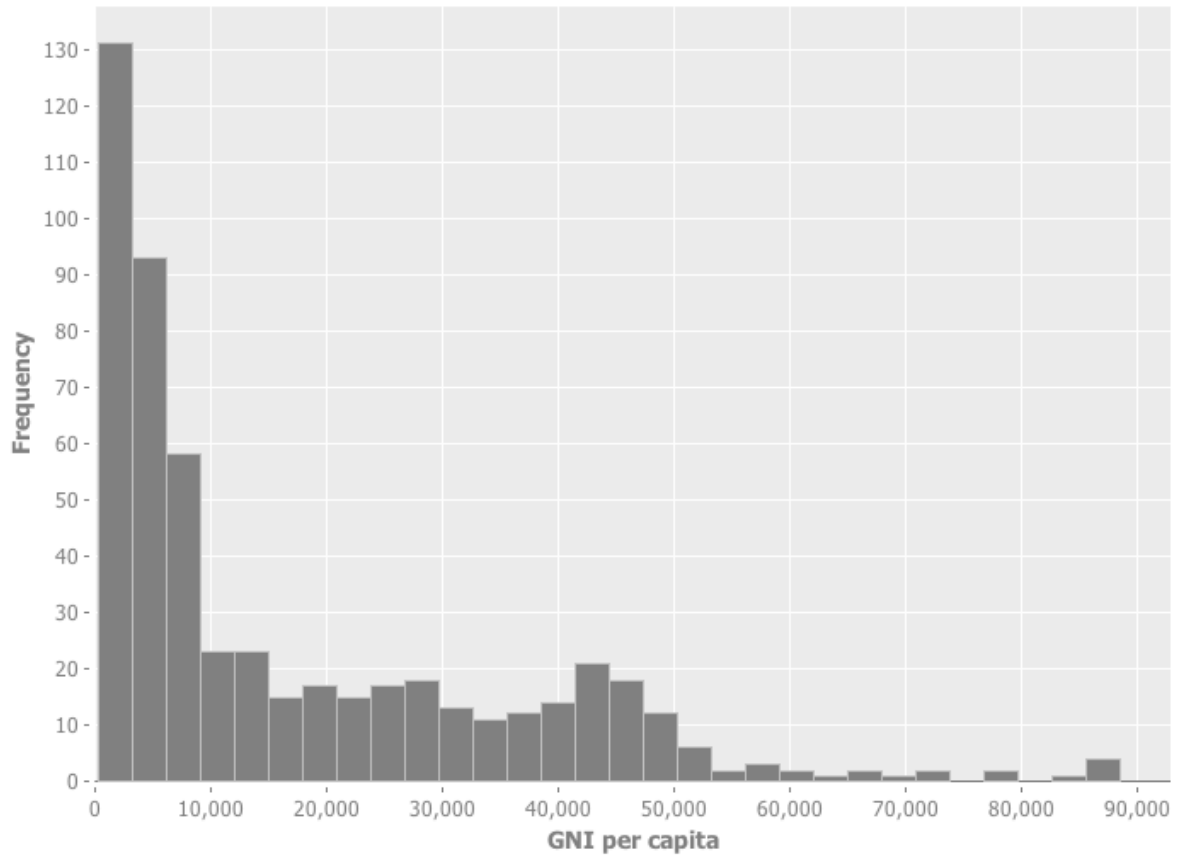


Incanter Plot

### Land Areas



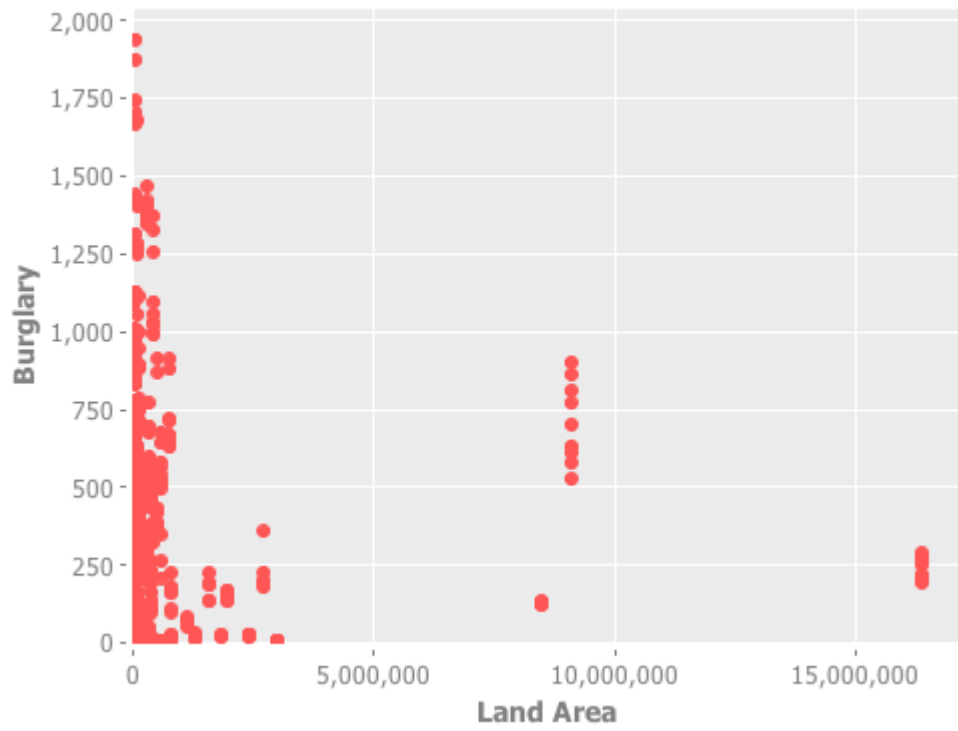
### GNI per capita



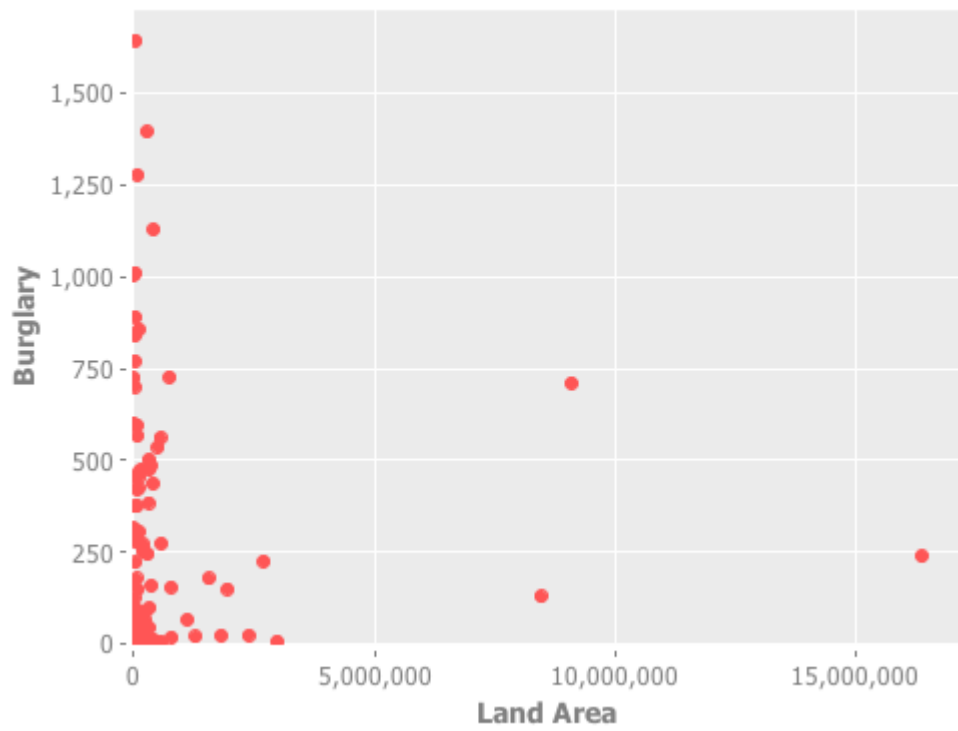
### Exponential distribution



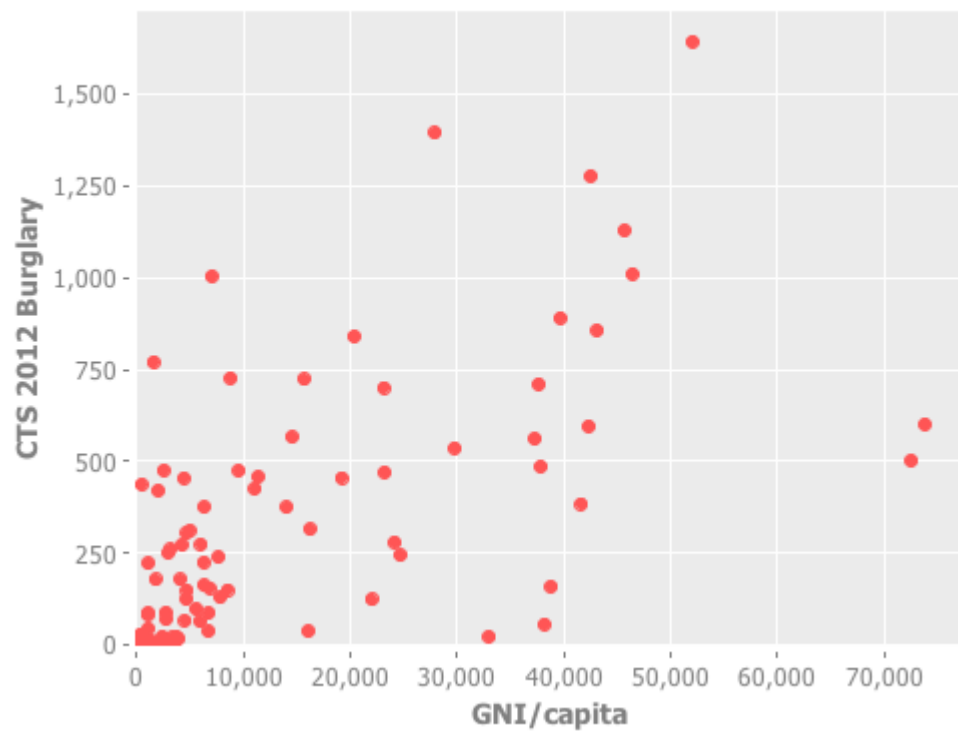
### Land Area and Burglary



### Land Area and Burglary



## GNI/capita and CTS 2012 Burglary

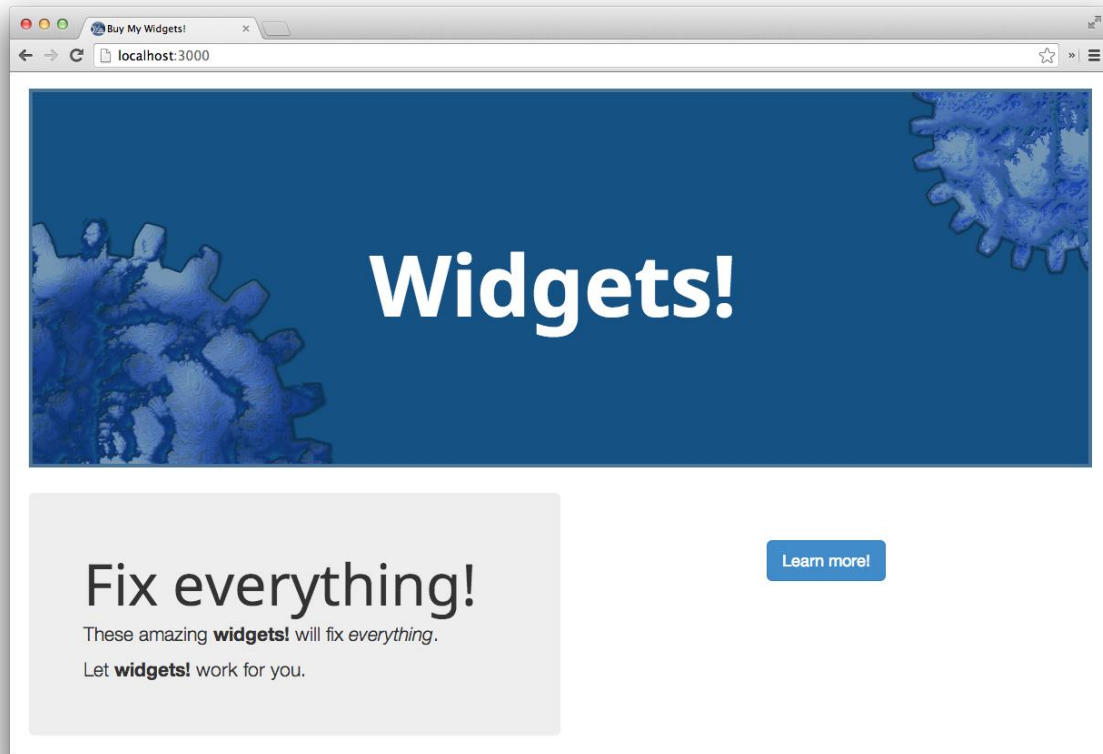


$$r_x = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

$$\sqrt{\frac{1}{(n-1)}}$$

## Chapter 8

# A/B Testing – Statistical Experiments for the Web



$$n = 16 \frac{\sigma^2}{\delta^2}$$



**Thank you!**  
For your order.



**Fix everything!**  
These amazing **widgets!** will fix *everything*.  
Let **widgets!** work for you.

[Learn more!](#)





Thank you!

For your order.

2426 observations.

ID	Group	Started At	Succeeded At
7352	:control	2013-12-01 17:06:00.407	2013-12-01 17:11:54.208
7349	:control	2013-11-29 01:17:17.825	none
6671	:control	2013-11-29 01:16:18.302	none
5536	:control	2013-11-29 01:15:22.965	none
7012	:test	2013-11-29 01:15:16.635	2013-11-29 01:15:42.119
6839	:test	2013-11-29 01:15:11.615	none
6051	:test	2013-11-29 01:14:33.155	none

# Not Done Yet

Group	Current Count
Control	306/1200 (25.5%)
Test	408/1200 (34.0%)

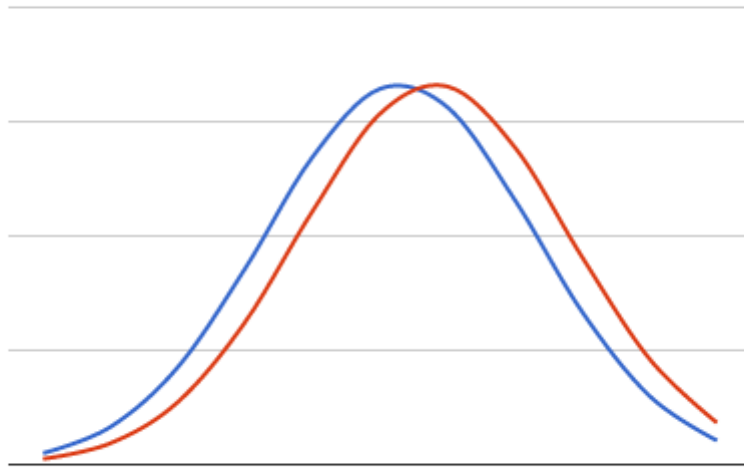
Back to the [list of observations](#).

# Results Grid

The test did better than the control.

	Success	Failure	Total
Control	252	974	1226
Test	317	883	1200

## Difference of Means



$$\mu = np$$

$$np(1-p)$$

$$SE(\bar{X}_t - \bar{X}_c) = \sqrt{\frac{var_t}{n_t} + \frac{var_c}{n_c}}$$

$$t = \frac{\bar{X}_t - \bar{X}_c}{SE}$$

$$v \approx \frac{\left(\frac{s_t^2}{N_t} + \frac{s_c^2}{N_c}\right)^2}{\frac{s_t^4}{N_t^2 v_t} + \frac{s_c^4}{N_c^2 v_c}}$$

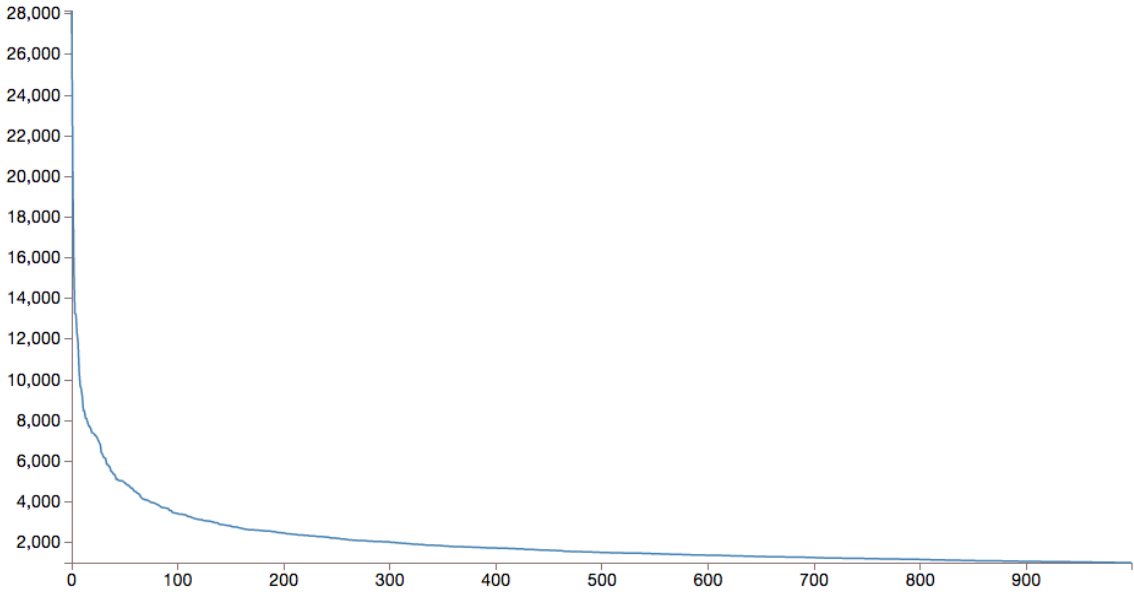
# Chapter 9

## Analyzing Social Data Participation

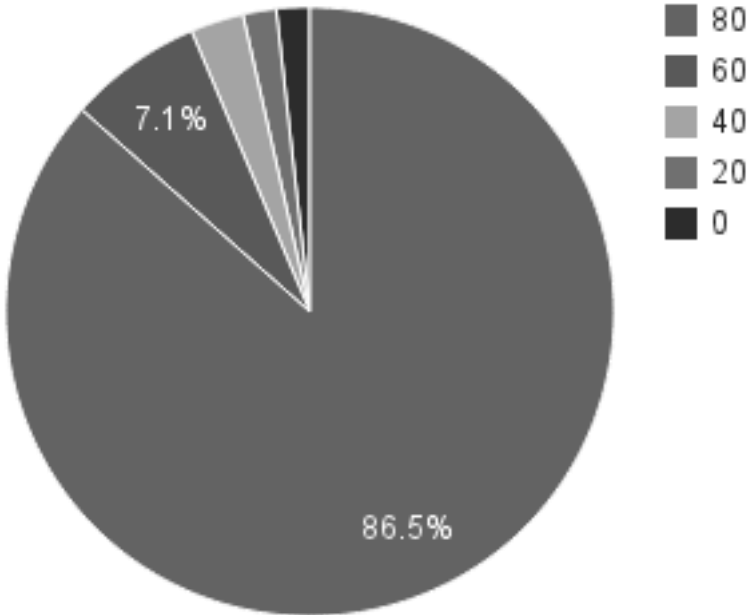
The screenshot shows the Stack Overflow website interface. At the top, there is a navigation bar with 'StackExchange' and links for 'sign up', 'log in', 'tour', 'help', and 'careers 2.0'. Below this is the 'stackoverflow' logo and navigation buttons for 'Questions', 'Tags', 'Tour', and 'Users', along with an 'Ask Question' button. The main content area features a 'Here's how it works' section with three icons: a question mark for 'Anybody can ask a question', two 'A' icons for 'Anybody can answer', and a large 'A' icon for 'The best answers are voted up and rise to the top'. Below this is the 'Top Questions' section with a filter for 'interesting' (459 results) and tabs for 'featured', 'hot', 'week', and 'month'. Two questions are listed: 'Windows File System MiniFilter Driver' (0 votes, 1 answer, 34s ago) and 'MVC Simple Membership - How to change login URL when unauthorized?' (0 votes, 2 answers, 46s ago). On the right, there is a 'CAREERS 2.0' section with job listings for ASP.NET Software Developers, iOS Developer, and Senior Android Developer.

stackapps.com.7z	<a href="#">4.9 MB</a>	
stackexchange_archive.torrent		
stackoverflow.com-Badges.7z	<a href="#">61.2 MB</a>	
stackoverflow.com-Comments.7z	<a href="#">1.4 GB</a>	
stackoverflow.com-PostHistory.7z	<a href="#">7.5 GB</a>	
stackoverflow.com-PostLinks.7z	<a href="#">19.7 MB</a>	
stackoverflow.com-Posts.7z	<a href="#">4.6 GB</a>	
stackoverflow.com-Users.7z	<a href="#">77.3 MB</a>	
stackoverflow.com-Votes.7z	<a href="#">307.0 MB</a>	
stats.stackexchange.com.7z	<a href="#">71.9 MB</a>	
superuser.com.7z	<a href="#">360.3 MB</a>	
sustainability.stackexchange.com.7z	<a href="#">1.5 MB</a>	

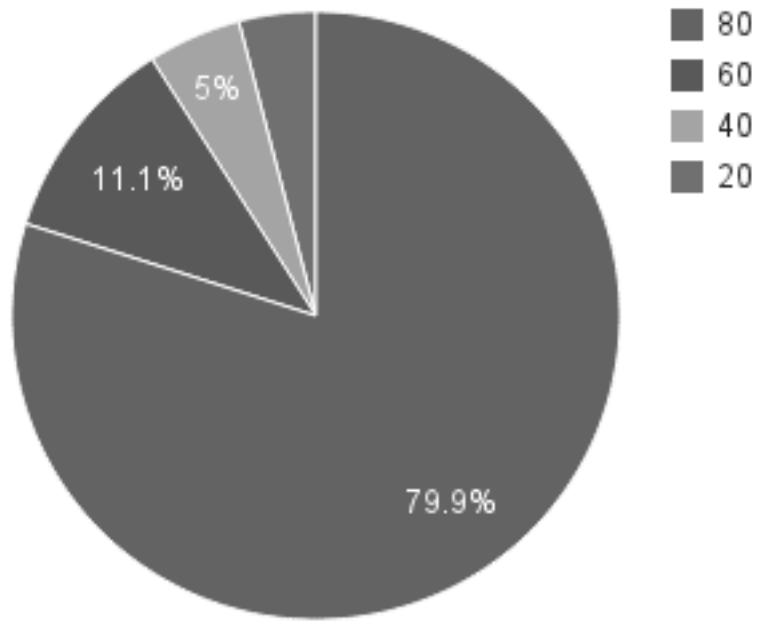
Post Counts by Rank



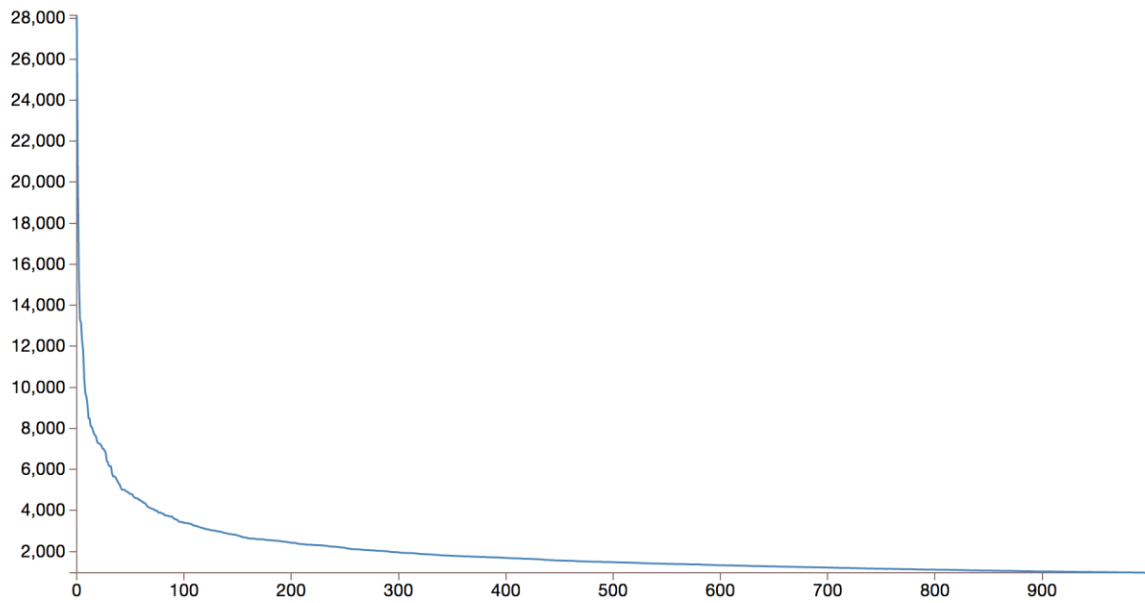
Contributor break-down, all posts



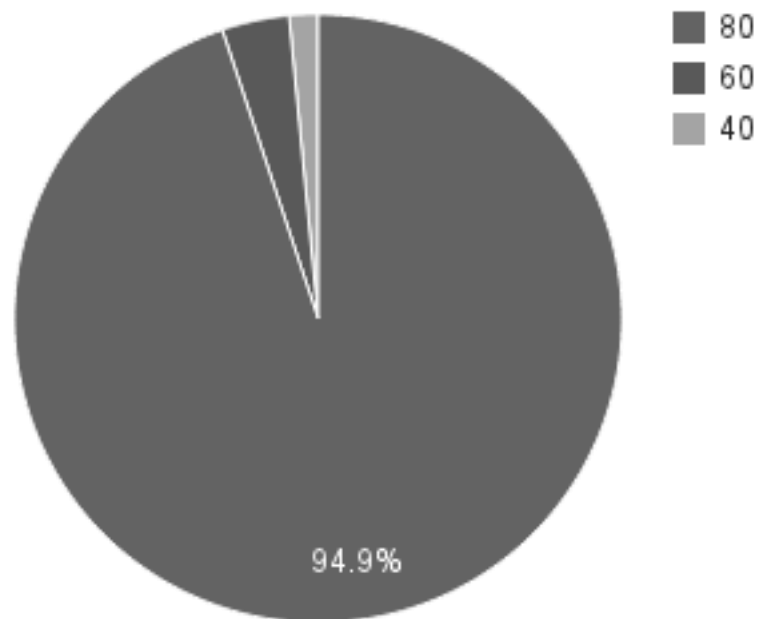
### Contributor break-down, questioners



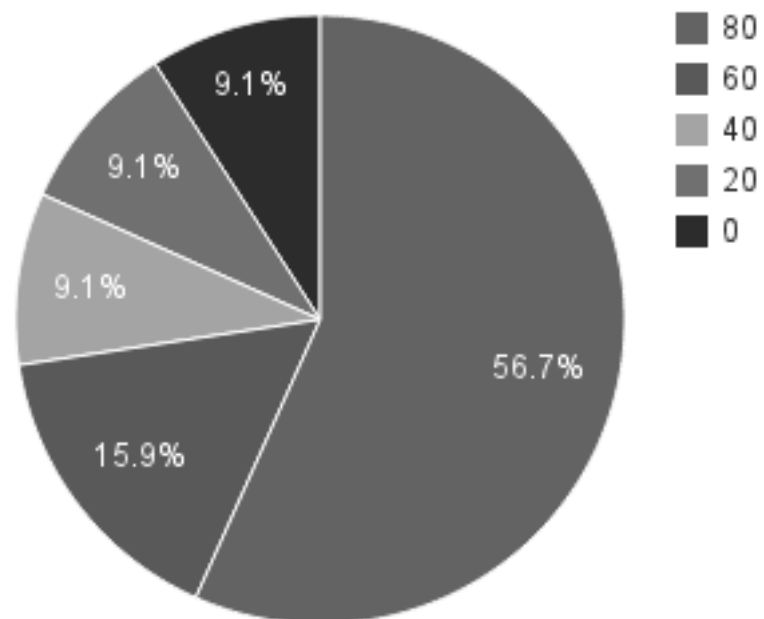
### Answer Counts by Rank



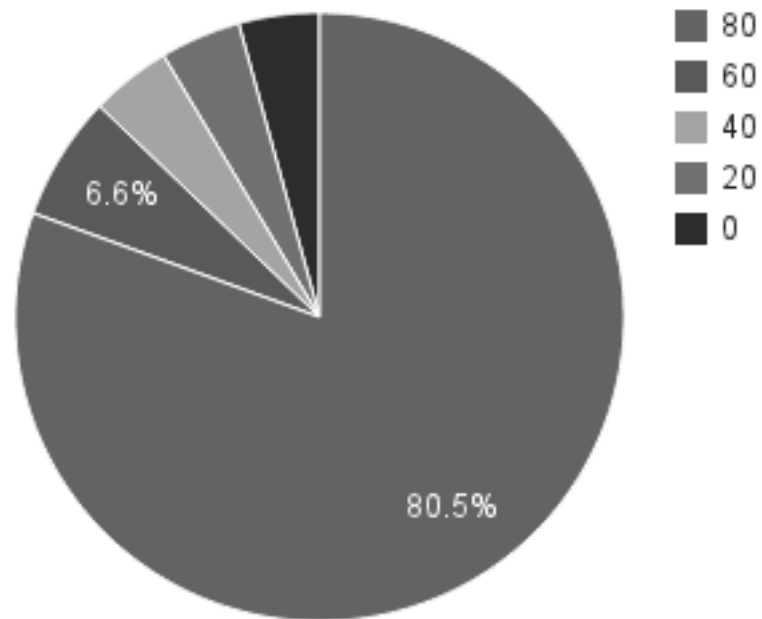
### Contributor break-down, answerers



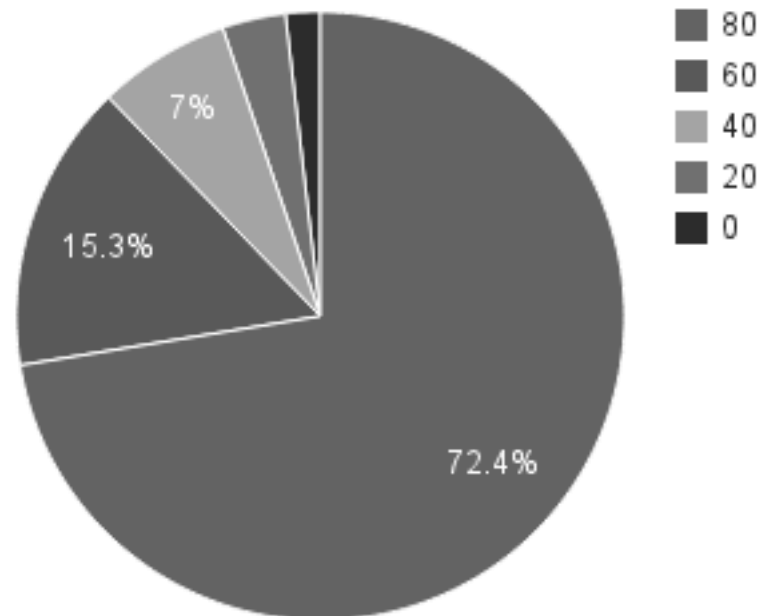
### Contributor break-down, only questions



### Contributor break-down, only answers

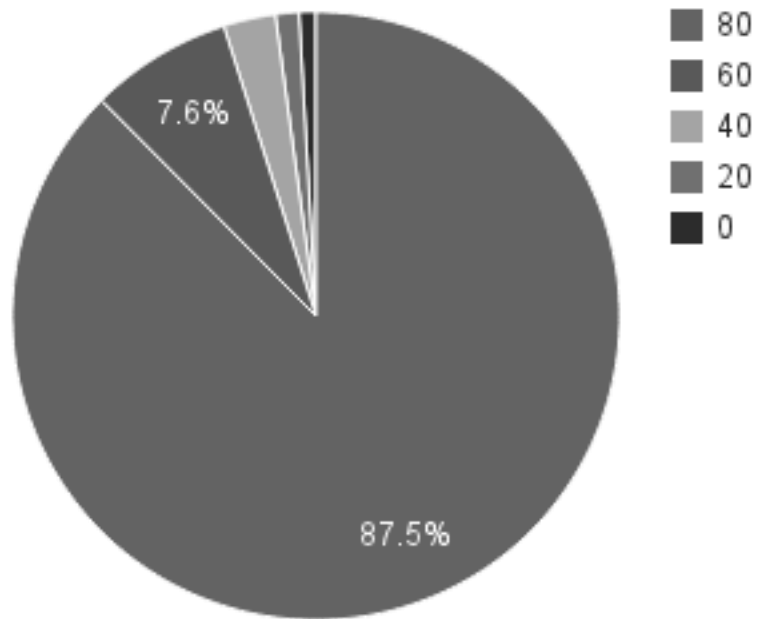


### Contributor break-down, both questions

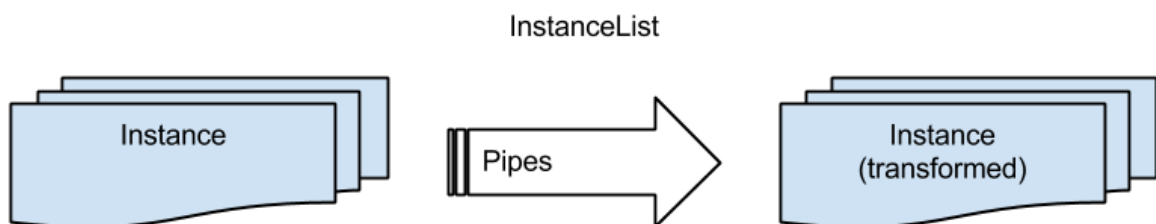
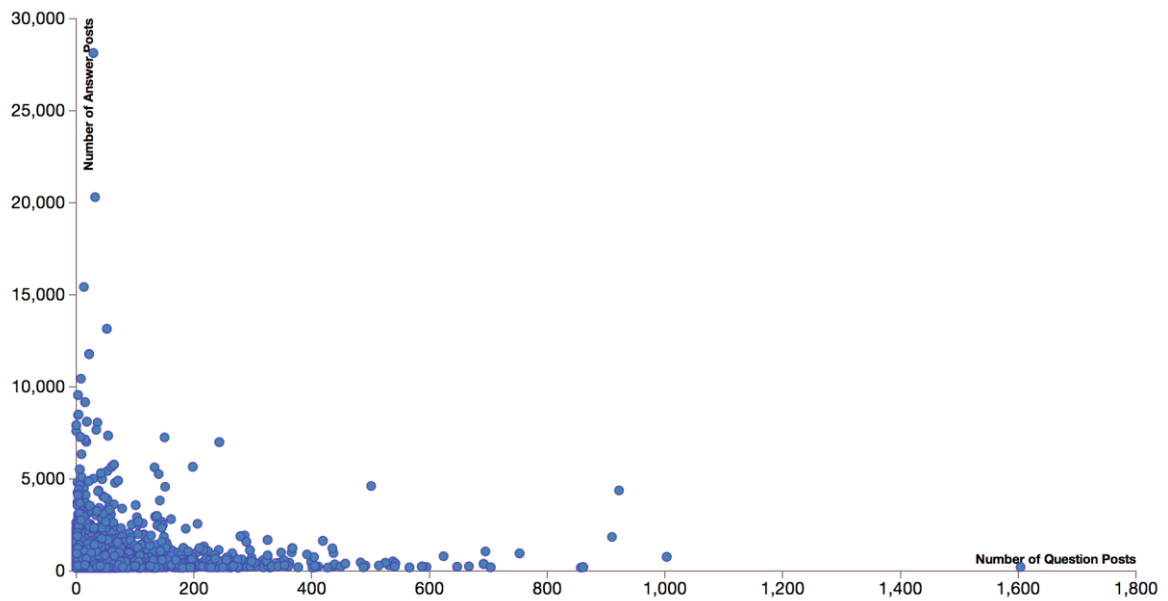




## Contributor break-down, both answers



## Answer Posts by Question Posts



# Chapter 10

## Modeling Stock Data

Daily prices Sep 1, 1995 - Jan 1, 2001 Update

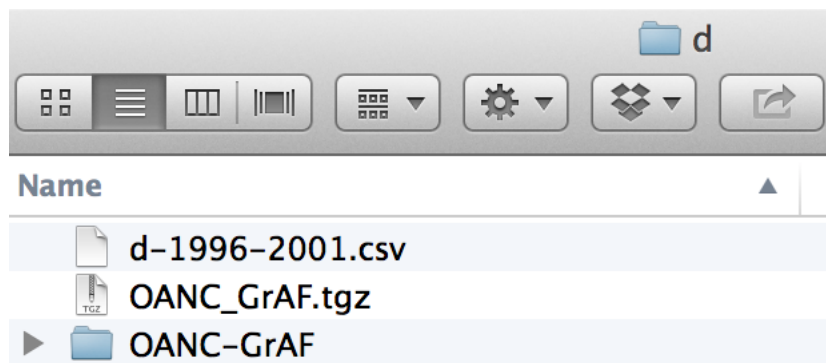
Date	Open	High	Low	Close	Volume
Dec 29, 2000	33.47	33.56	33.09	33.50	857,800
Dec 28, 2000	33.62	33.62	32.94	33.47	961,200
Dec 27, 2000	33.50	33.97	33.19	33.56	992,400
Dec 26, 2000	32.88	33.69	32.88	33.62	660,600
Dec 22, 2000	32.25	32.78	32.06	32.69	938,800
Dec 21, 2000	32.19	32.50	31.75	32.34	1,577,400
Dec 20, 2000	31.16	32.12	30.94	31.88	1,888,600
Dec 19, 2000	31.44	31.53	30.94	31.03	1,466,800
Dec 18, 2000	30.88	31.25	30.59	31.03	1,242,200
Dec 15, 2000	30.50	30.91	30.47	30.75	2,317,800
Dec 14, 2000	30.88	30.88	30.38	30.50	1,613,000
Dec 13, 2000	30.09	30.94	30.06	30.81	1,795,000
Dec 12, 2000	30.28	30.41	30.00	30.09	1,029,400
Dec 11, 2000	30.09	30.12	29.59	30.03	1,377,400

Historical chart

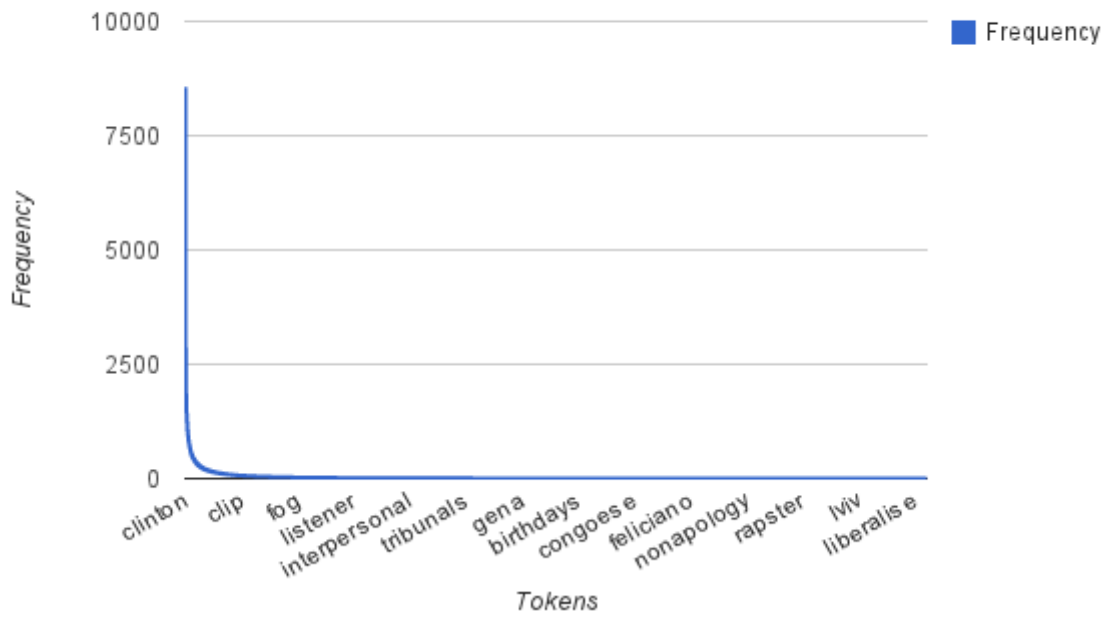


Export

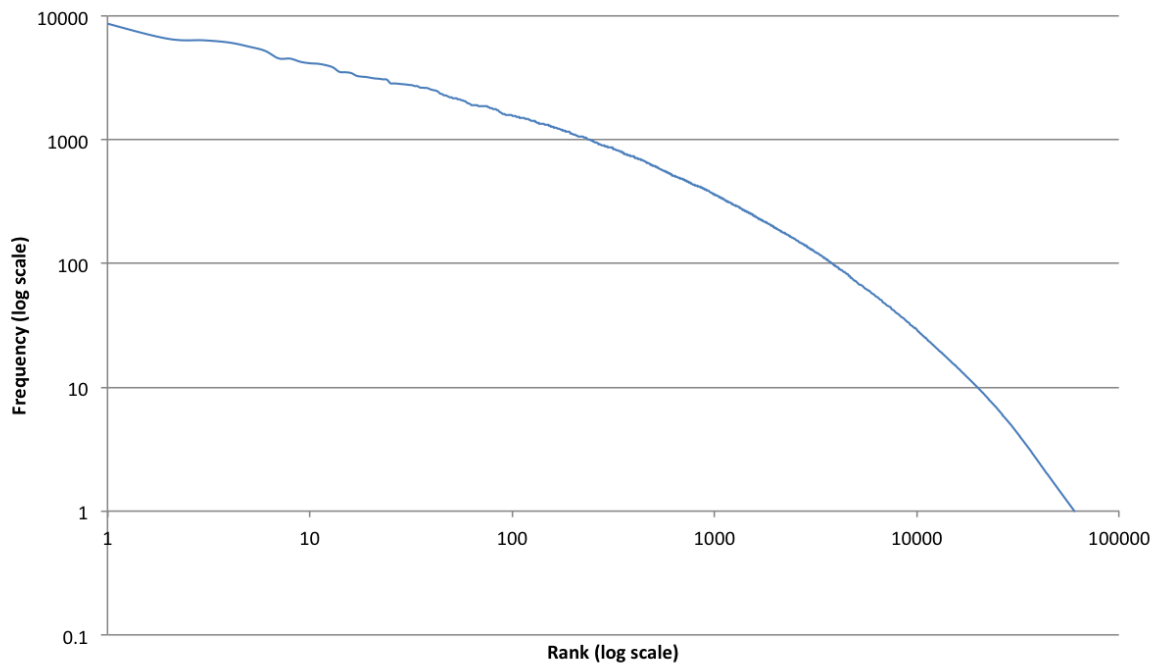
[Download to spreadsheet](#)

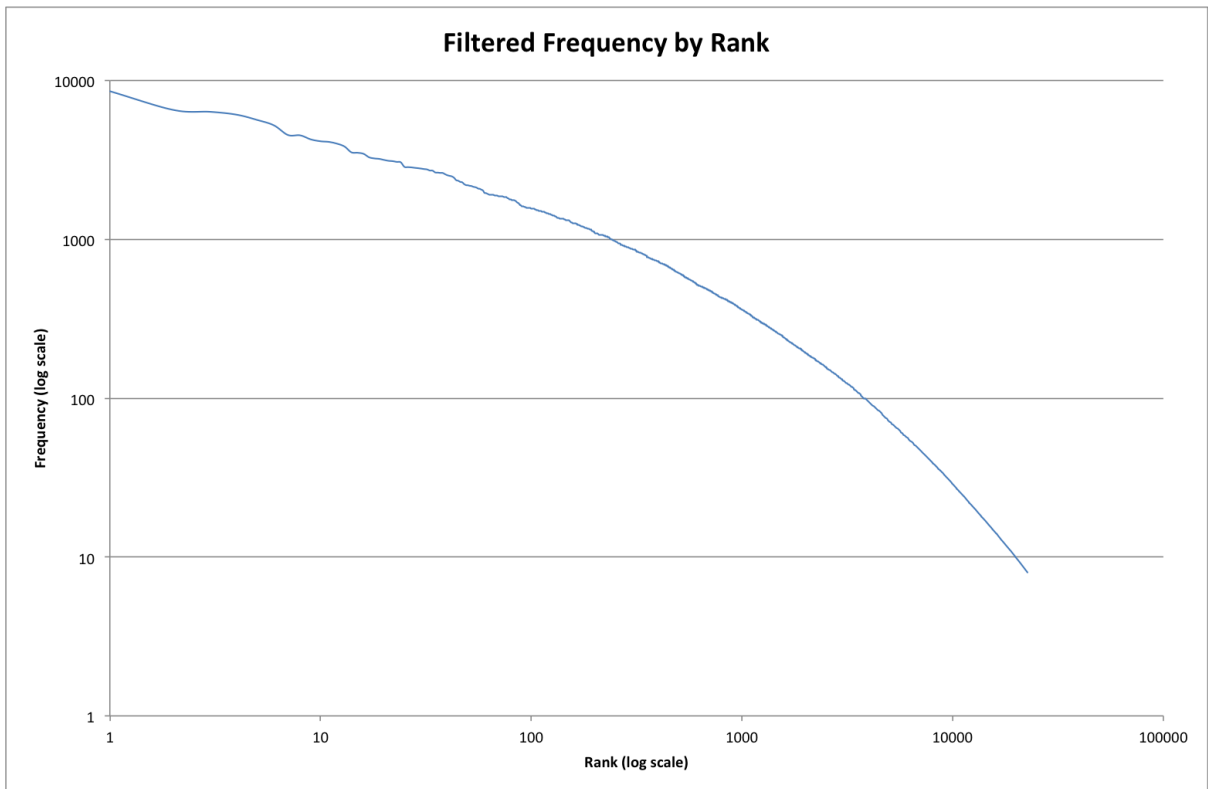
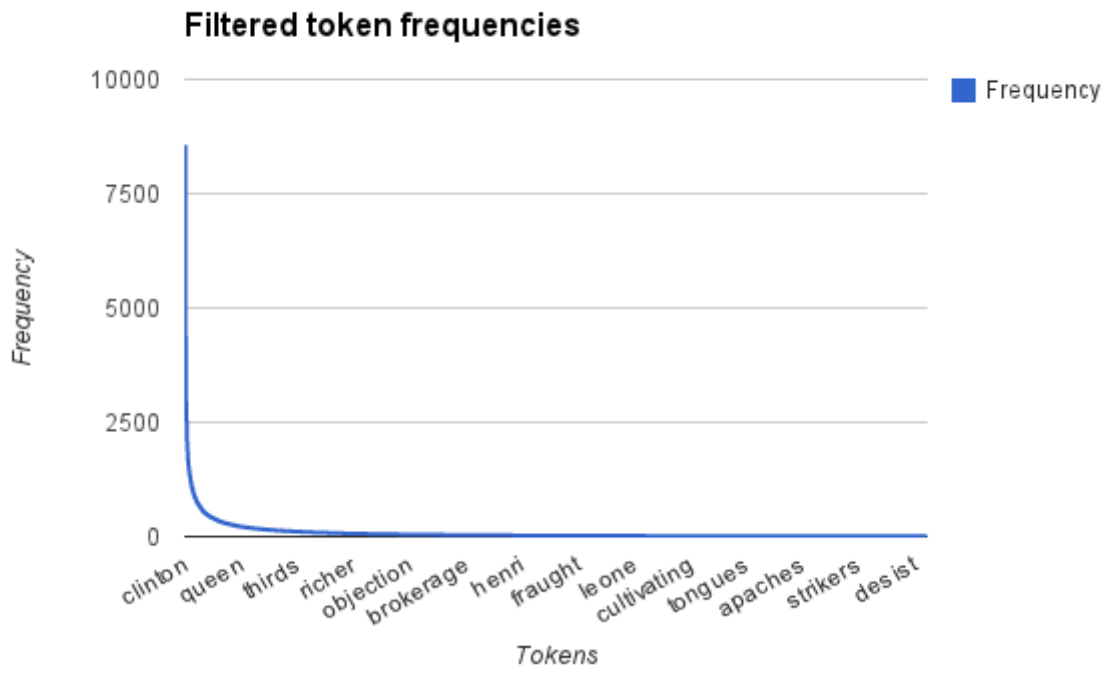


### Token frequencies



### Token Frequency by Rank





**Closing Price of Dominion, Inc, 1996-2000**

