DATA VISUALIZATION TYPES

OVERVIEW

As the purpose of this document is to help guide information designers to 1) use a shared vocabulary when discussing concepts related to data visualization 2) choose the right visualization type to represent a dataset and 3) cue the designer into some of the critical decisions that need to be made regarding that visualization, I have divided this documentation into four sections.

First, I'll be introducing the basic concepts involved in choosing the right visualization type: data types, modes of tasks, types of tasks, and types of displays. This will include some info on how best to optimize the display itself.

Next, I'll walk through some of the more common ways to visualize data like bar charts, scatter plots, and time series graphs. I will evaluate each one of these visualization types by a number of different considerations or desiderata (in attached excel spreadsheet).

Finally, I'll document a taxonomy of common visualization types and their use cases.

WHY VISUALIZE

People use data to make decisions. The expressed purpose of almost every information designer is to present data in ways that will help people make decisions. In almost every case, this means the data will be presented in some graphical form over simply being listed in a table.

At every step of the information design process, it's important to come back to this important concept – let the design aid the decision maker.

Tufte breaks down this concept of 'making decisions easier' into four basic principles:

1. Show the data
2. The viewer should think about the data, not about the design or methodology (unless part of the users analytical task)
3. Utilize space efficiently
4. Access several levels of detail

Show the data – don't hide or obscure the data with a complex design or hidden options. Unless the expressed purpose of the viewer, don't encourage them to think about 'how the graph is designed' or 'how to do it differently' – the data itself should take center stage.

Utilizing space effectively means maximizing what Tufte calls the ink-to-info ratio – there should not be any excess, unnecessary information on a graph that doesn't help the viewer make her decision. Finally, accessing several levels of detail means giving the user the ability to drill down/look up from anywhere in the graph (e.g., hovering over a bar in a bar graph to display it's value).

Each section to follow in the documentation will all address how to present data to conform to the users decision-making process and goals. But all will presuppose adherence to these four basic principles.

DATA TYPES

Categorical Data: text describing the nature of the data (e.g., age)
Quantitative Data: numerical data to represent a category (e.g., 25 years)

All graphs have categories – which cue the viewer into how the data is being divided. Consider a bar chart showing quarterly sales growth in each region across the country. The overall category describing the nature of the data 's presentation is 'quarterly' performance, with each quarter being broken down into different regions.

Categorical data can be classified at one of three levels:

1) Nominal – there is no intrinsic order to the categories (e.g., accounting, sales, marketing). It does not make sense to say that 'accounting' comes before 'sales.'
2) Ordinal – there is an intrinsic order to the categories, but the distance between them may vary/there is no index (e.g., first place, second place, etc.). First naturally comes before second, but it does not always make sense to say, for instance, that first place finisher had half the time of the second place finisher.
3) Interval/Ratio – interval scales have an intrinsic order, and differences between them are meaningful, meaning each incremental value has the same 'weight.' When a graph uses interval data as a category, it is always the result of transforming a continuous, quantitative scale into a categorical one. Take age – it does make sense to say that someone who is 2 y/o is twice as old as someone who is 1 y/o.

However, displaying every value from a continuous scale on a graph's axis can make the data difficult to understand, so these values are usually 'binned' - grouped together into distinct clusters, e.g., years 0 - 10, 11 - 20, 21 - 30 – and so on. This data type may or may not be indexed.

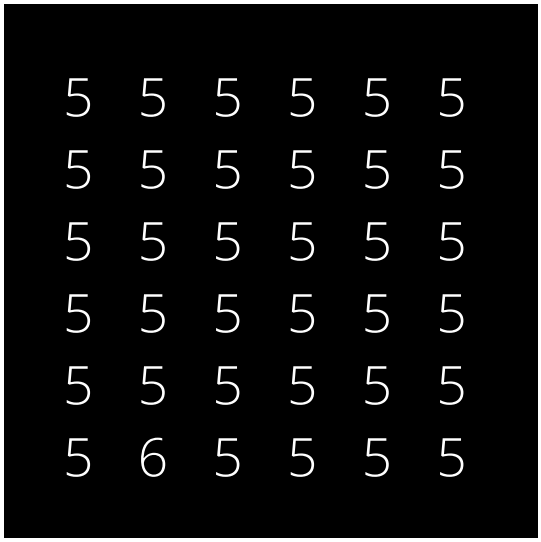
Depending on the scale of the data, different visualizations will need to be used. More on this later.

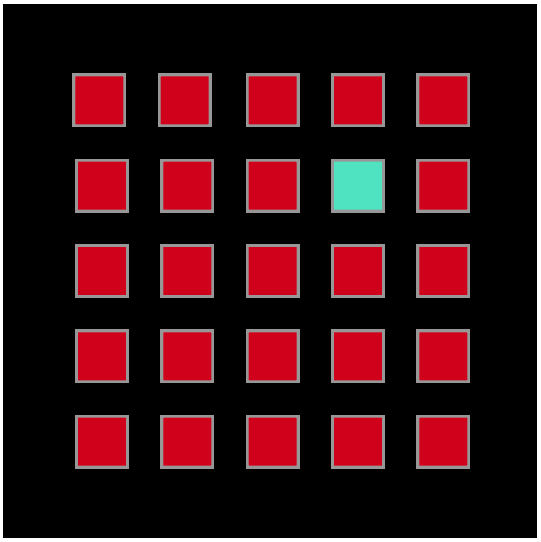
VISUAL PERCEPTION AND PREATTENTIVE ATTRIBUTES

Briefly, humans can process information in the environment using two different methods:

1) Parallel
2) Serial

Parallel, preattentive processing occurs when a person unconsciously takes in multiple pieces of information from the environment all at the same time, while serial processing is conscious deliberate intake of information one item at a time. Consider the following graphics – in which is it easier to detect the outlier?





When considering the plot on the left, you likely had to search serially, one number at a time, before you could find the '6'. However, you could point out the outlier on the right before you even consciously considered it. The former is an example of serial search, the latter – preattentive parallel.

Like the example, humans are better at unconsciously processing some attributes (like color) over others (fine differences between lines of the same general shape). The goal of every data visualization designer is to leverage preattentively processed attributes so the user doesn't have to expend as much cognitive energy to understand the display. The preattentive attributes that we have the easiest time processing, in descending order, are:

1. Position along a common scale
2. Position along identical, nonaligned scales
3. Length
4. Angle
5. Area
6. Volume
7. Color hue
8. Color saturation
9. Density

After which come a few verbal and auditory cues.

To complicate matters, a user's expectations of her own task can influence her ability to easily detect data in a display – so in addition to leveraging preattentive attributes, it's also good practice to ensure that the user knows precisely what results she is looking for, especially when designing more complex displays.

Takeaway: In order to reduce complexity, leverage the preattentive attributes most easily processed by the perceptual system. This reinforces Tufte's first principle of 'showing' the data.

ANALYTICAL PATTERNS AND TASKS

Viewers leverage these preattentive attributes to 'take in' a graphical data display, but a task like "find the only blue square" is much simpler than a real business decision like "which department is over budget this year, by how much, and what should we do about it?" How is this gap bridged?

After preattentive processing, we complete what Tufte calls 'Analytical Tasks' – tasks that are more complex than simply processing preattentive attributes but don't capture all of the complexity of the decision space. These analytical tasks serve as intermediary steps that a viewer needs to accomplish in order to make a real business decision. They'll become clearer with examples:

- Identify Similarities and Differences
- Identify Outliers
- Identify Patterns and Trends
  - Are they steady? Fluctuating? Cyclical? Seasonal?
- Identify Relationships, Causality, and Directionality
- Identify Steep and Shallow Distributions
- Identify Concentration of Data Points/Gaps between Data Points
- Identify Geospatial Location of Data Points
- Identify Overlapping Data Distributions
- Identify Transitions between Distributions
- Identify Natural Ranking of Data
- Identify the Shape of Data
- Identify Part to Whole Relationships
- Lookup a Raw Value

Completing these analytical tasks is foundational to interpreting a display.

As such, it's important to identify any intermediary tasks the user has to accomplish when considering how to visualize a dataset. For instance, a viewer's ability to make a decision about an outlier might be hindered when using a histogram (which affords processing the data's shape) but facilitated when using a box and whisker plot (which affords processing shape and unusual values).

DESIGNING FOR ANALYTICAL TASKS

Notice the first two distinct preattentive attributes in the list: position and length. Few identifies these as the most effective preattentive attributes to leverage when communicating quantitative data. Information designers should use these attributes when helping their viewers complete analytical tasks.

In a graphical display, 'position and length' will be encoded as 'points,' 'lines' and 'bars.' Each attribute has it's own strengths and weaknesses for communicating information.

POINTS

Points emphasize the individual values of the data they represent. Like a table, they can be easily used for lookup whether in a Cartesian plot or on a geographical map. In most cases, points lack the visual weight to efficiently communicate the data's 'shape,' so they will rarely be used alone (with scatter plots/any graph that only displays quantitative information as a common exception). Points leverage the 'position along a common scale' attribute when displayed on a plot, and when enough are present they can say a lot about a data's distribution. They also help the user identify: outliers, patterns, relationships (in a scatter plot), and overlap.

LINES

Lines are used to 1) connect points for easier visual processing, 2) display trends within data and 3) display a data's shape. Connecting points with a line encourages the viewer to focus on patterns rather than individual points. This makes it a good candidate for identifying: overlap, trends, patterns and deviation from patterns.

BARS

Bars give visual weight to the values of things. The large area of the bar compared to a line or a point can cue the viewer in to smaller differences between values. Bars, however, are not good for visualizing trends. This makes them useful for identifying: similarities and differences, values, and ranking.

Each type of display - like a probability distribution or a time series - lends itself to using specific preattentive attributes to help the viewer complete specific analytical tasks. Next I'll exhaust the different types of displays, which analytical tasks they help complete, and how to leverage the corresponding preattentive attributes.

TYPES OF DISPLAYS

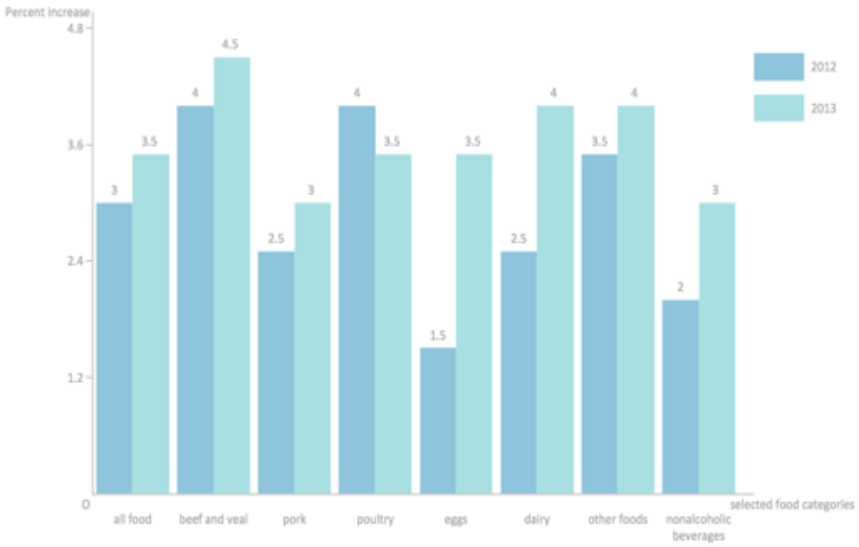
Considering the last three sections together – data types, preattentive attributes and their visual encodings, and analytical tasks – has implications for how certain types of graphs should be displayed. The major subcategories of displays are: 1 Dimensional, 3 Dimensional, Multi-Dimensional, Temporal, Geospatial, Tree, and Network. Within Multi-Dimensional there are six major types: nominal comparison, ranking, part to whole, deviation, frequency, and correlation. I'll discuss each one of these display types, its visual encoding, data types, and tips to make its use as seamless as possible for the viewer, so she can focus her energy on making decisions rather than struggling to understand the data.

MULTI-DIMENSIONAL (n-D)

Items with n attributes become points in an n dimensional space. For instance, a simple bar chart plots a category's value in 2D space using height to represent values. It is good practice to keep the primary visualization 2D with additional dimensions like time or tick size being represented by small multiples or toggle buttons, rather than making the display appear 3D. Adding a third dimension to multidimensional display (like a 3D bar chart) usually only serves to clutter the display and make it more difficult for the viewer to focus on the data itself, so use 3D sparingly and only when absolutely necessary. For this reason, it will not be dealt with at length in the documentation.

NOMINAL COMPARISON

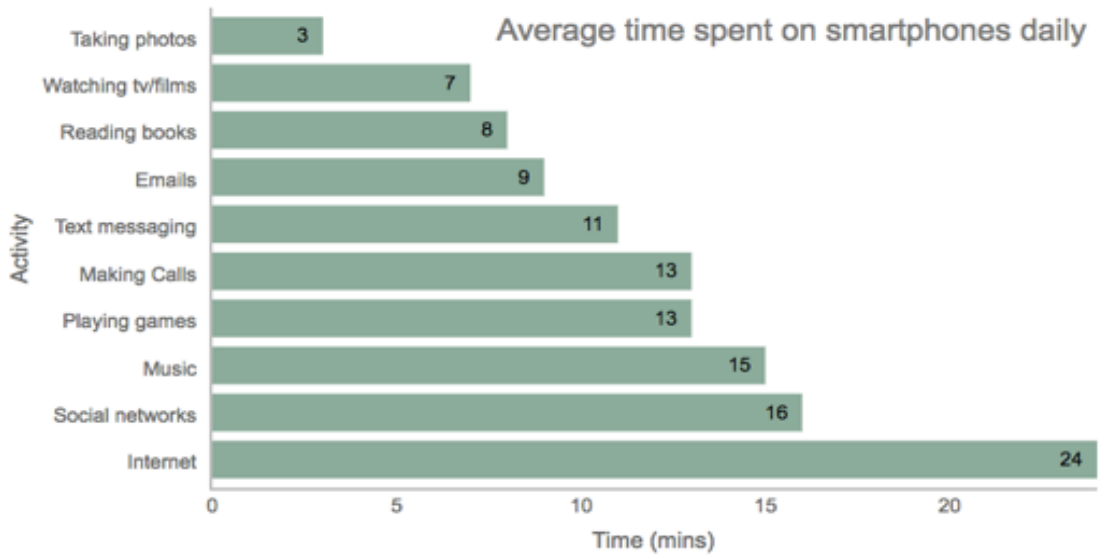
Example: display the percentage increase of various food purchases over two years.



Nominal comparisons offer simple comparisons of categorical subdivisions in no particular order. They should be used when the data type is **nominal** and the analytical tasks include **ranking and value lookup**. Nominal Comparisons should always be encoded as **bars**, since bars give appropriate visual weight to the values of each subcategory.

RANKING

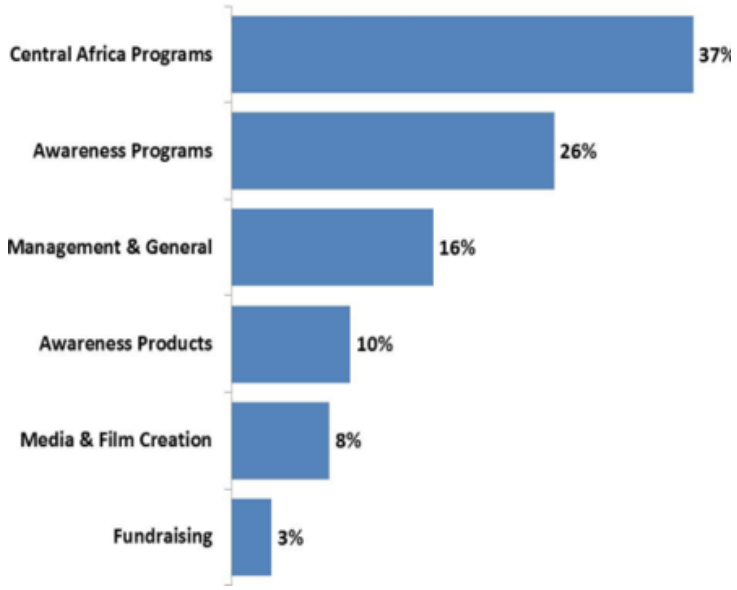
Example: Ranking different smartphone activities by how much time people spend doing them.



Ranking is very similar to nominal comparison, with a few slight modifications. First, the data on the axis that represents the independent variable can be either **nominal or ordinal**, and the analytic task **ranking** should be prioritized over **value lookup**. Finally, the categories should be ordered by their measurement of the dependent variable, either ascending or descending. Visual encoding should continue to be a **bar**.

PART TO WHOLE

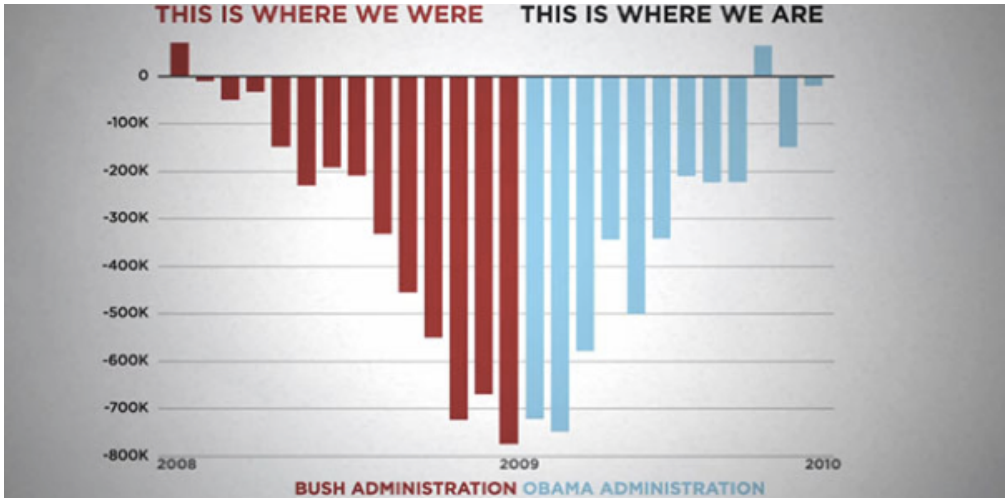
Example: display the revenue generated by each marketing program as a proportion of total revenue.



Part-to-Whole displays show categorical subdivision's measurements as ratios of the whole category measurement. This type of display should be represented by **bars**, as they effectively utilize the preattentive attributes area, volume and length to communicate relative values. The data type can be **nominal or ordinal** (part to whole relationships for interval/ratio scales are best represented as histograms), and the analytical tasks of the user are identifying **Part to Whole relationships** and **Ranking** values.

DEVIATION

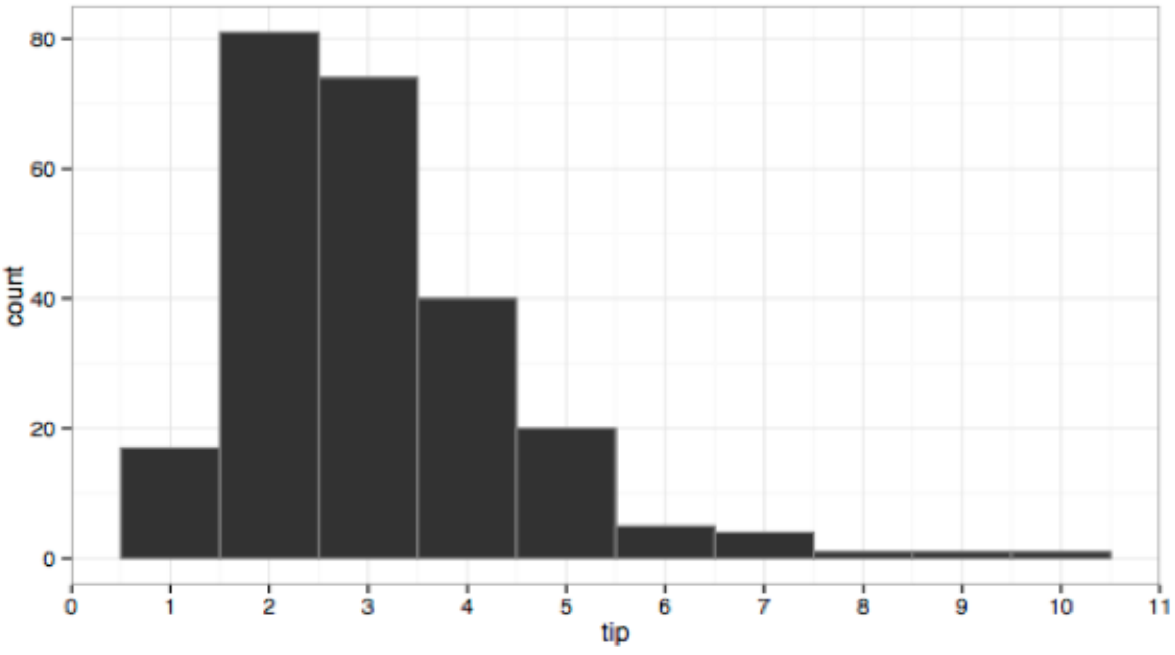
Example: display cumulative job gain/loss statistics by president



In deviation displays, increases and decreases in subcategories are compared against a reference line (in the above example, the reference line is the number of jobs at the beginning of the last year of the bush administration). Another common example might be revenue and expenditures graphed along the same reference line to represent profit. Deviation displays can be encoded in a few different ways. If the viewer's goal is to **identify a pattern**, use only **lines**. If the viewer needs to attend to patterns as well as **individual values**, use **points connected by lines**. Finally, if the viewer is only attending to the **values**, use the visually dense **bars**. All three scales of data (**nominal, interval, ordinal/ratio**) can be displayed in a deviation display. Note that if the visualization is displaying changes over time, the horizontal axis should always display time to conform to viewers' existing knowledge about time series graphs.

FREQUENCY DISTRIBUTION

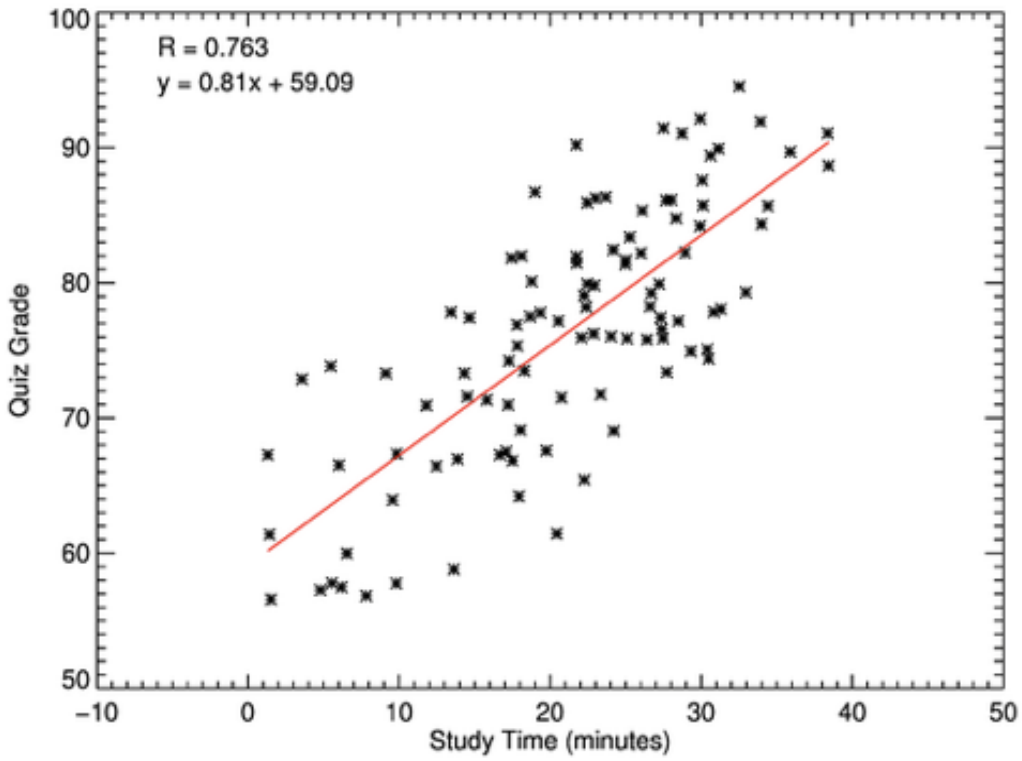
Example: display the number of times customers tip a certain dollar amount from 1 to 10.



Frequency distributions display the number of observations of a variable at multiple intervals. Unlike nominal comparison or ranking displays, frequency distributions display a continuous, **interval/ratio** variable on the independent axis, usually in ranges or 'bins.' The number of times an observation falls within the range of that bin, the corresponding bar is incremented. Analytical tasks that lend themselves to frequency distributions include **identifying part to whole relationships, patterns/ trends**, as well as assessing the overall '**shape**' of the data. **Bars** should be always be used over lines, though a reference line displaying a smoothed density distribution can also be included in specific instances.

CORRELATION

Example: What is the relationship between time spent studying for a test and the grade received on that test.



Correlation matrices are used to determine if the variability in one measurement is related to the variability in another measurement. They are most often used to predict one value from another – how well am I likely to do on this exam if I study 20min? 40min? The analytical tasks of the user that warrant using a correlation matrix include identifying **outliers, patterns and trends, relationships and causality, concentration** of data in a certain area, and value **lookup**. Since each measurement situated by two **interval/ratio** scale variables, it can only be encoded as a point. A **line** of best fit should almost always be used to orient the viewer to the shape of the data and make on-the-fly predictions.

TEMPORAL

Temporal displays capture the measurement of one or more variables over time. If the purpose of the display is to overview a series of discrete events, **bars** should be used to emphasize the individual values (duration) of those events (e.g., displaying a timeline or Gantt chart, step charts excluded). If the purpose is to identify a **pattern** of behavior over time of a continuous variable, **lines** should be used to emphasize the overall pattern (e.g., displaying stock price). The most common examples of temporal displays are time series graphs, gantt charts, and timelines.

GEOSPATIAL

Geospatial displays compare data across a map. The simplest displays are dot distribution maps, which show a single **point** on a map to represent the occurrence of an event at that location (e.g., show the distribution of citizens with cardiovascular problems near a hydrofracking site). In other cases, different preattentive attributes are used to communicate patterns like **volume** (cartograms) and **color saturation** (choropleth). In almost all cases, the analytical tasks of the viewers are **Patterns/Trends; Outliers; Location; and Concentration**.

TREE

Tree displays use identical **lines** and **points** to represent a hierarchy of categories and subcategories, ending in terminal nodes that represent individual data points. If all lines and nodes are the same length and size, the viewer can assess **System Complexity, Concentration, and Part to Whole** relationships.

NETWORK

At its most fundamental level, a network display shows interrelationships between entities in a system. Those relationships may or may not have direction. This type of display is different from hierarchical trees in that most network diagrams are flat – either that individual entities/nodes and their relationships all coexist at the same categorical level, or that the relationships describe something other than hierarchy. Network displays utilize lines and points, to help the viewer accomplish the analytical tasks: identify **relationships**, understand **system complexity**, and identify **concentration** of nodes within a system.

ONE-DIMENSIONAL

There is only one, one-dimensional data display: a list. Lists are rarely visualized. They are most often sorted alphabetically, or give the user the ability to choose the sorting technique. List data is almost always nominal.

THREE-DIMENSIONAL

In most cases, three-dimensional displays are used to display three dimensional, spatial attributes of a real world object that cannot be seen by the naked eye. Computer graphics of molecules or star systems are common examples. Volume rendering is another common scientific use case – most 3D brain imaging use Volume Time Course data to display changes in brain activity over time using many 2D image slices.

MODE OF TASK

There is one final constraint on choosing the proper visualization type – task mode. Task modes describe 'how,' or the manner in which, one uses a display to complete any of the analytical tasks mentioned above. Understanding a potential user's task mode can help information designers create a display that further eases the decision making process. There are five common types of task mode:

LOOKUP

Simply, lookup displays are used to lookup facts. Data for lookup tasks are usually presented in tables, or have 'details on demand' features embedded within the display. Information is usually communicated through verbal channels (words and numbers), not graphical ones. Many visualization experts find that lookup is used too often in the business intelligence space. The goal of visualization is to use graphics to reveal trends and help decision makers. Rarely does a tabular representation expand beyond simply displaying the data. If a user's only analytical task is lookup, then it is appropriate to use a table to display data. The table should facilitate the lookup process by providing the user with categories and criteria by which to sort and filter.

NARRATIVE

Narrative displays offer text descriptions and explanations next to visualizations. They answer questions like “what is the current status of the data?” and “how did we get here?” The most common method narrative displays use to make sense of data is the story – communicated both through text and graphics. Infographics fall into this category. There are two major design considerations of narrative displays. First, the design must closely correspond to the mental model the viewer has of the process being detailed. Since the information takes on narrative form, the story in the display must share many characteristics with the story inside the viewers mind. Second, the display is most often static, dense and requires a lot of

MONITORING

Monitoring displays help users maintain ongoing awareness of what’s going on. They are typically dynamic, allow for direct manipulation of the graphical interface, and are seldom used unless the viewer needs to take prompt action. A viewer’s ability to make a prompt decision will increase as the complexity of the display decreases; displays should increase the info/ink ratio as much as possible, clear out any chart junk, and heavily utilize the most easily processed preattentive attributes. Any added complexity will weigh too heavily on the cognitive system, pulling resources away from the display and the decision. For example, an airplane cockpit with a monitoring display will be more effective one with a denser narrative display. Like narrative displays, monitoring displays should be closely aligned with the mental models of the users.

EXPLORATORY DATA ANALYSIS (EDA)

EDA has two goals: 1) explore data to find facts, and 2) make sense of those facts. The display user likely will not know the best visualization type/statistical method to process the data, and will need to try multiple before finding the most useful one. To make this task easier, Few suggests all EDA interfaces should include:

1. Rapid and fluid interaction with data
2. The ability to compare multiple datasets
3. The ability to view the same dataset from multiple perspectives
4. Access to basic statistical functions
5. Efficient combination of data from different sources

PREDICTION

Predictive displays are used to visualize graphs like probability distributions and scatter plots, as well as provide some inferential statistical functions like what-if/ predictive analysis. Most statisticians will do this without any visualization (i.e., a process runs in the background and outputs a number or two), however this method doesn't encourage the use to think about the method behind the analysis or its results. Depending on the context this can be a good or a bad thing, but special attention needs to be paid to the role of the end user in predictive analysis.

WHAT'S NEXT

A few general comments about method:

A direct manipulation display is any graphical display that allows the user to change the interface in real time. The most common cases of direct manipulation are 1) hover to get more details of something on the graph, 2) zoom in/out or change one or both of the scales, and 3) update the query parameters.

Every direct manipulation display should try to conform as closely as possible to Shneiderman's "Visual Information Seeking Mantra." The mantra provides a general guide about which information in a display should be available 'at first,' which information should be available after the user 'zooms' and which information should the user be able to change. In short the answer is 'all of it' but the mantra outlines the general steps a user might take to dive deeper into a dataset.

1) Overview: gain an overview of the entire collection
2) Zoom: zoom in on items of interest
3) Filter: filter out uninteresting items
4) Details on demand: Select an item or group and get details when needed
5) Relate: view relationships among items
6) Extract: allow extraction of sub-collections and of the query parameters

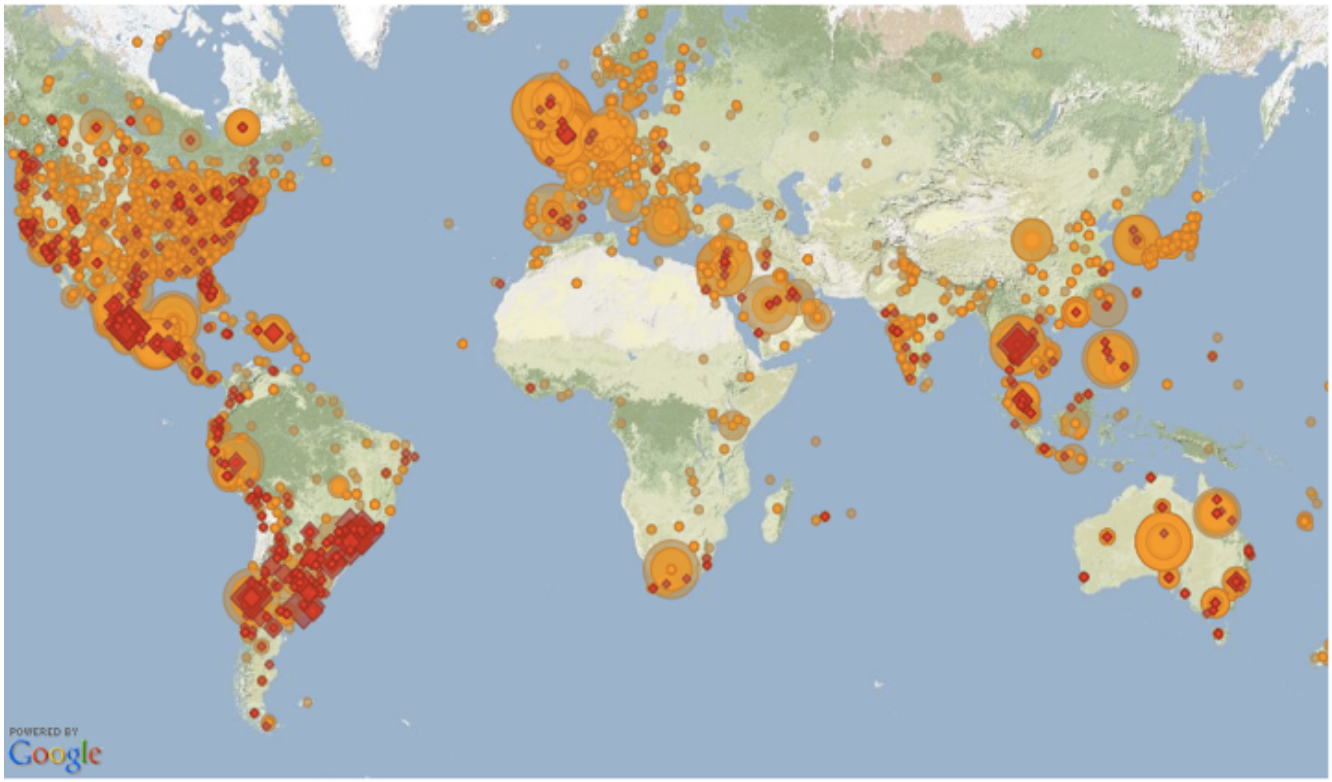
Holding off on any commentary about Data State Model until I have more information about the systems in use.

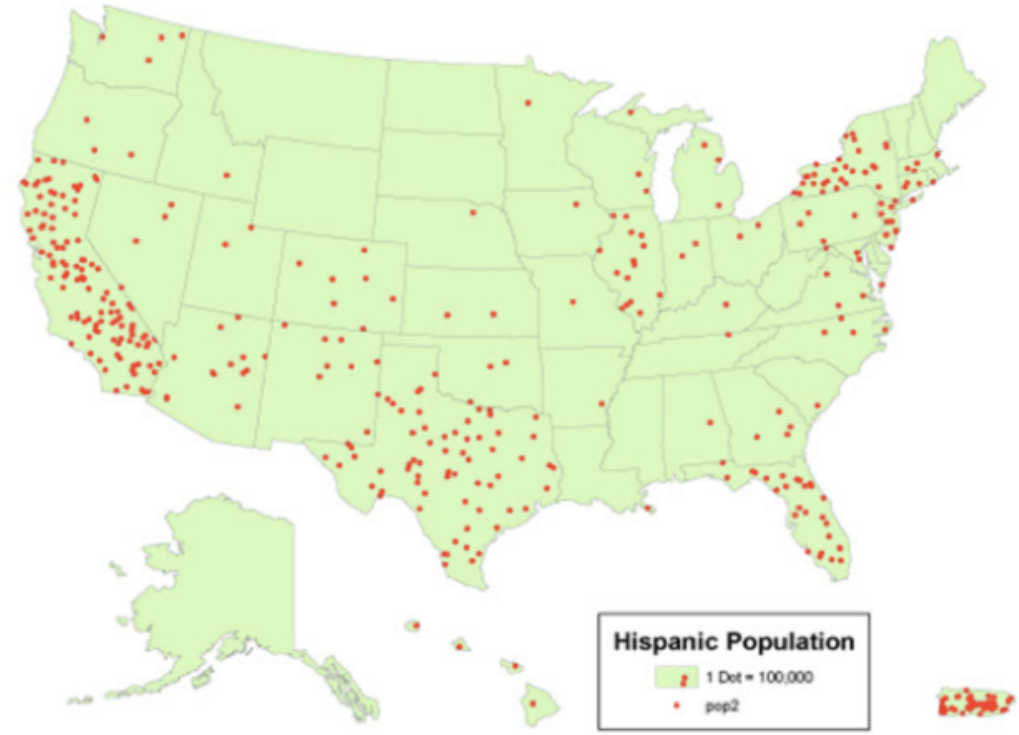
IMAGES

LIST

No visualization.

DOT DISTRIBUTION MAP




Hispanic Population
1 Dot = 100,000
pop2

TIME SERIES/STEP CHART

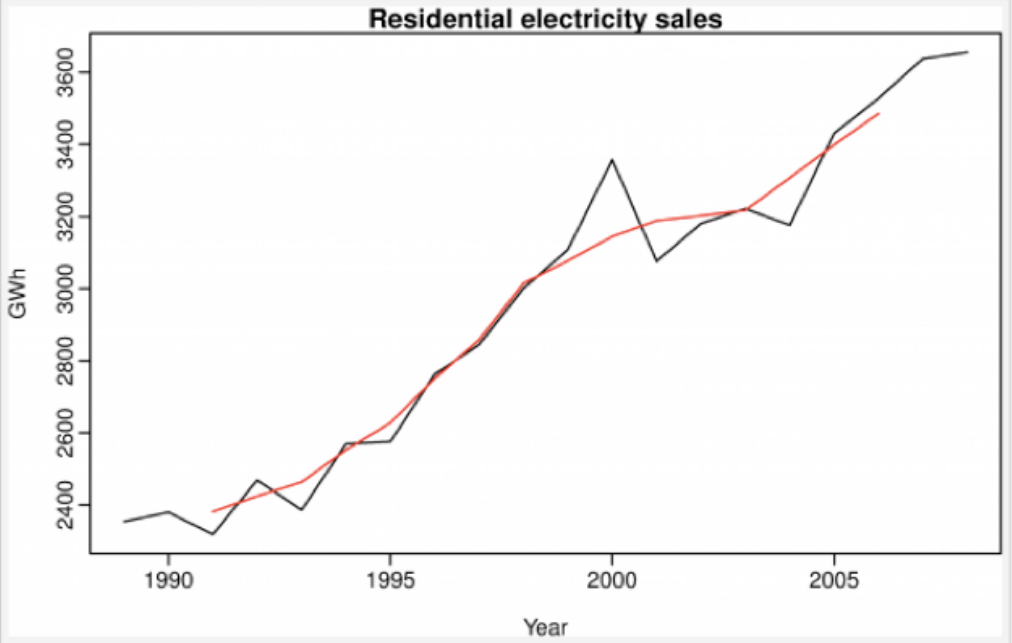
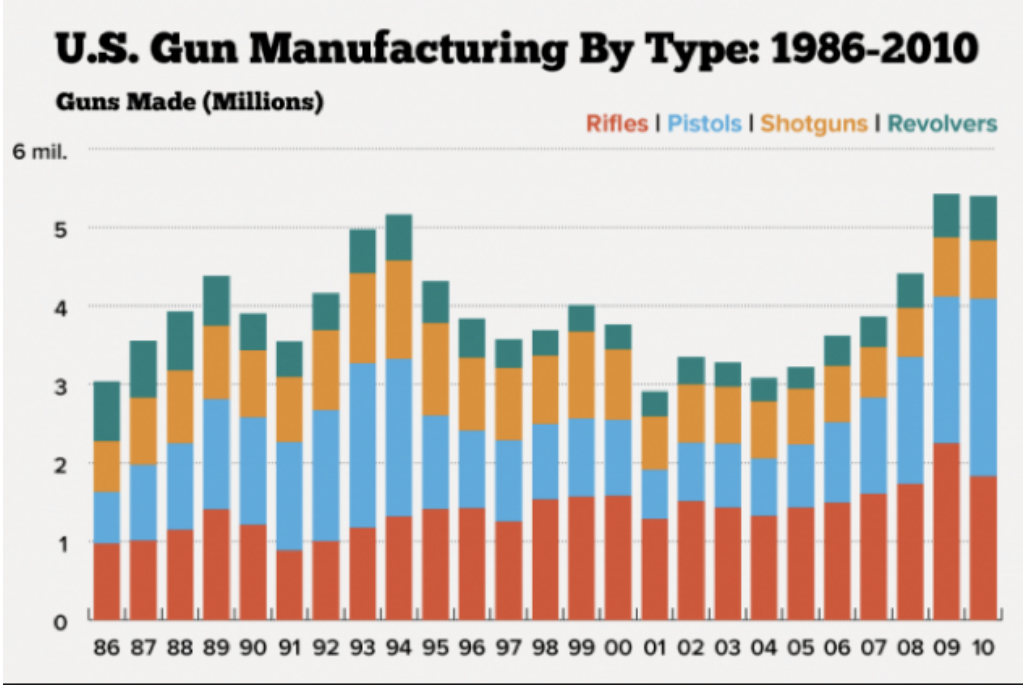
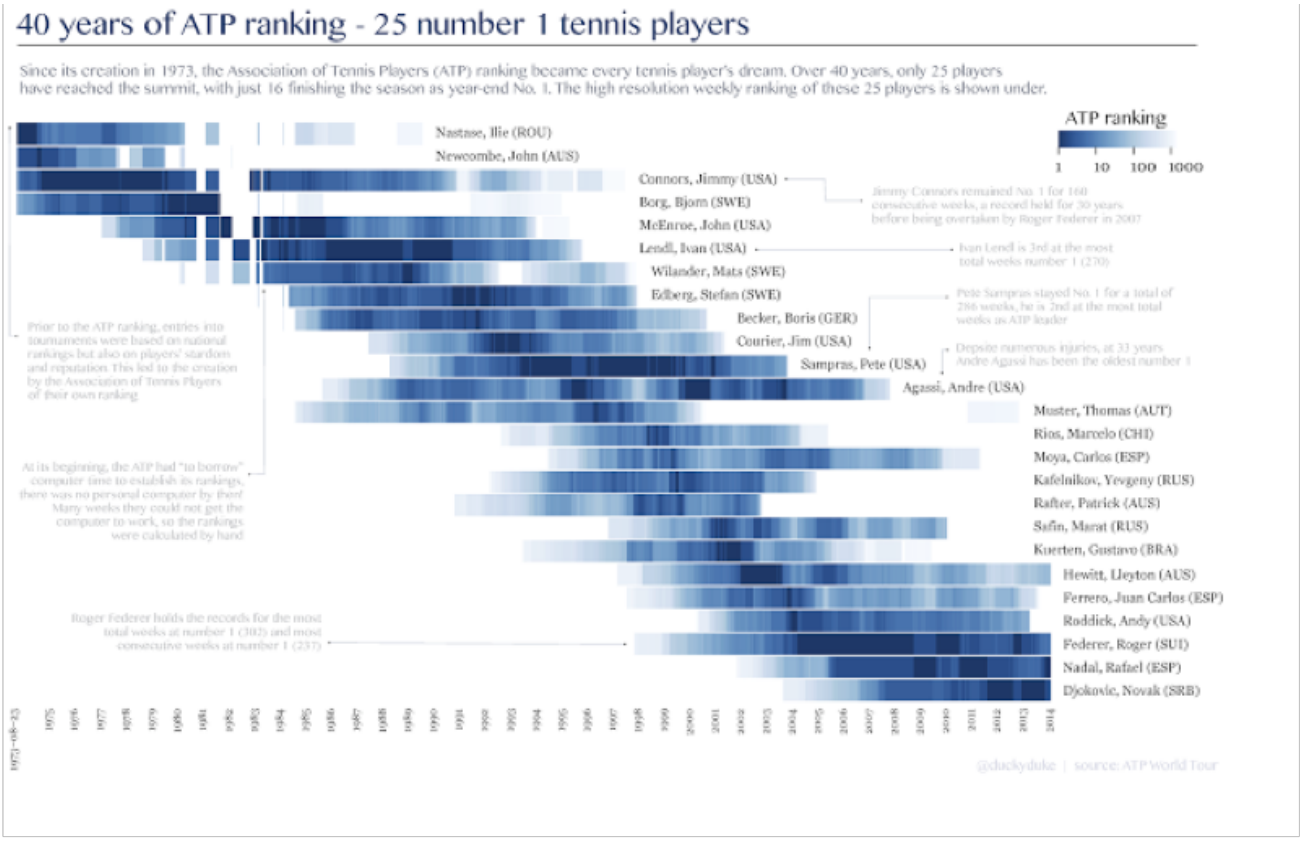


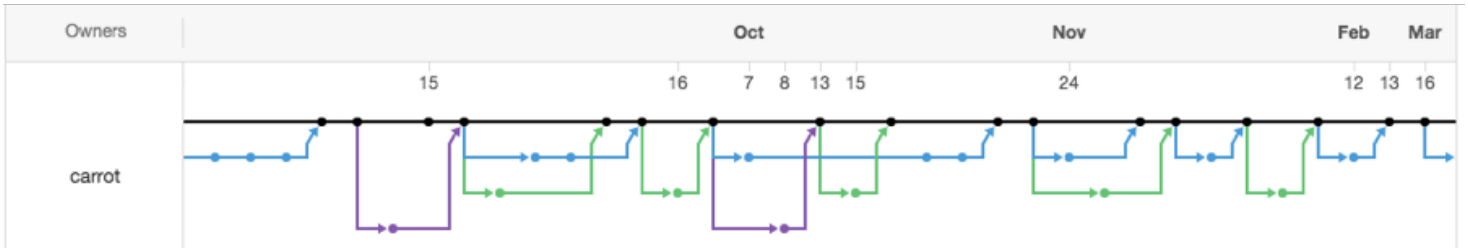
Figure 6.7: Residential electricity sales (black) along with the 5-MA estimate of the trend-cycle (red).

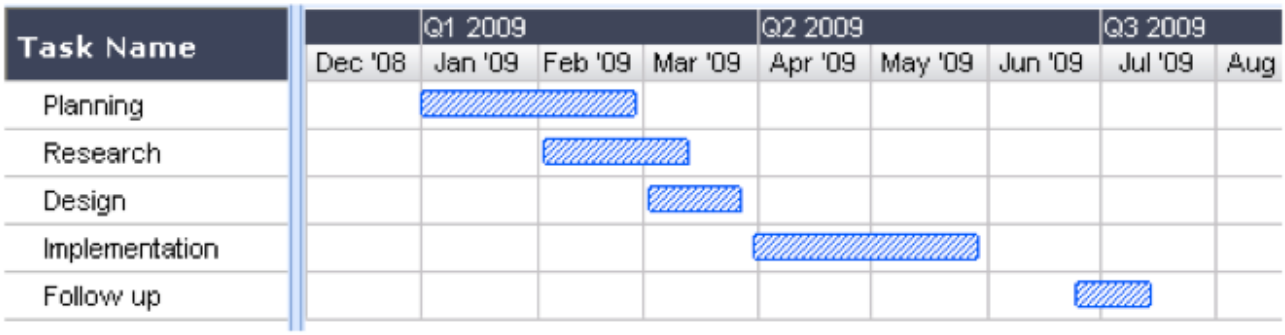



Edging Higher... and Higher
U.S. stocks rise to records
Normalized As Of 11/23/2015
S&P 500 Index
Dow Jones Industrial Average
NASDAQ Composite Index
Russell 2000 Index
112.02
110
105.348
95
90
85
Nov Dec
2015
Jan
Feb
Mar
Apr
May
Jun
2016
Jul
Aug
Sep
Oct
Nov
Bloomberg

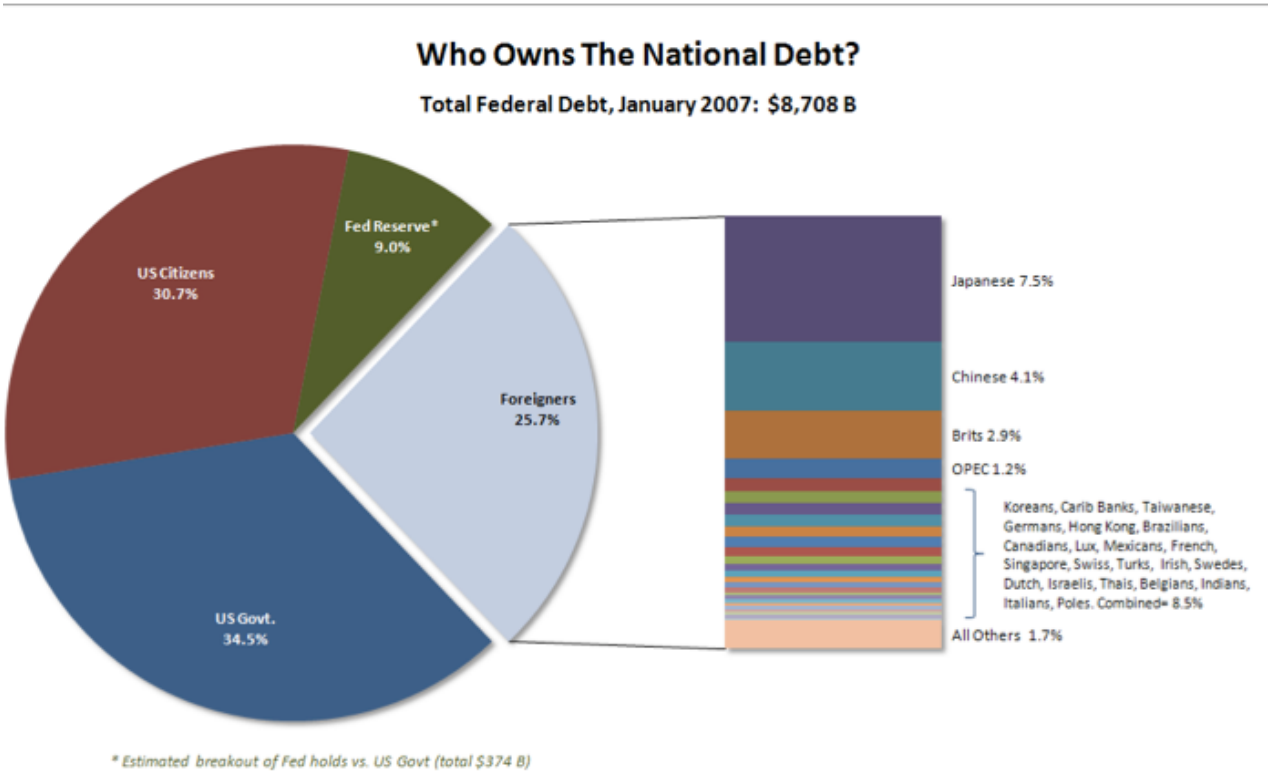
GANTT CHART


40 years of ATP ranking - 25 number 1 tennis players
Since its creation in 1973, the Association of Tennis Players (ATP) ranking became every tennis player's dream. Over 40 years, only 25 players have reached the summit, with just 16 finishing the season as year-end No. 1. The high resolution weekly ranking of these 25 players is shown under.
ATP ranking
1
10
100
1000
Nastase, Ilie (ROU)
Newcombe, John (AUS)
Connors, Jimmy (USA)
Borg, Bjorn (SWE)
McEnroe, John (USA)
Lendl, Ivan (USA)
Wilander, Mats (SWE)
Edberg, Stefan (SWE)
Becker, Boris (GER)
Courier, Jim (USA)
Sampras, Pete (USA)
Agassi, Andre (USA)
Muster, Thomas (AUT)
Rios, Marcelo (CHI)
Moya, Carlos (ESP)
Kafelnikov, Yevgeny (RUS)
Rafter, Patrick (AUS)
Safin, Marat (RUS)
Kuerten, Gustavo (BRA)
Hewitt, Lleyton (AUS)
Ferrero, Juan Carlos (ESP)
Roddick, Andy (USA)
Federer, Roger (SUI)
Nadal, Rafael (ESP)
Djokovic, Novak (SRB)
Jimmy Connors remained No. 1 for 160 consecutive weeks, a record held for 30 years before being overtaken by Roger Federer in 2007
Ivan Lendl is 3rd at the most total weeks number 1 (270)
Pete Sampras stayed No. 1 for a total of 286 weeks, he is 2nd at the most total weeks as ATP leader
Despite numerous injuries, at 33 years Andre Agassi has been the oldest number 1
Prior to the ATP ranking, entries into tournaments were based on national rankings but also on players' stardom and reputation. This led to the creation by the Association of Tennis Players of their own ranking
At its beginning, the ATP had "no borrow" computer time to establish its rankings, there was no personal computer by then! Many weeks they could not get the computer to work, so the rankings were calculated by hand
Roger Federer holds the records for the most total weeks at number 1 (302) and most consecutive weeks at number 1 (237)
1973-08-23
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
@duckyduke | source: ATP World Tour


Owners
Oct
Nov
Feb
Mar
15
16
7
8
13
15
24
12
13
16
carrot

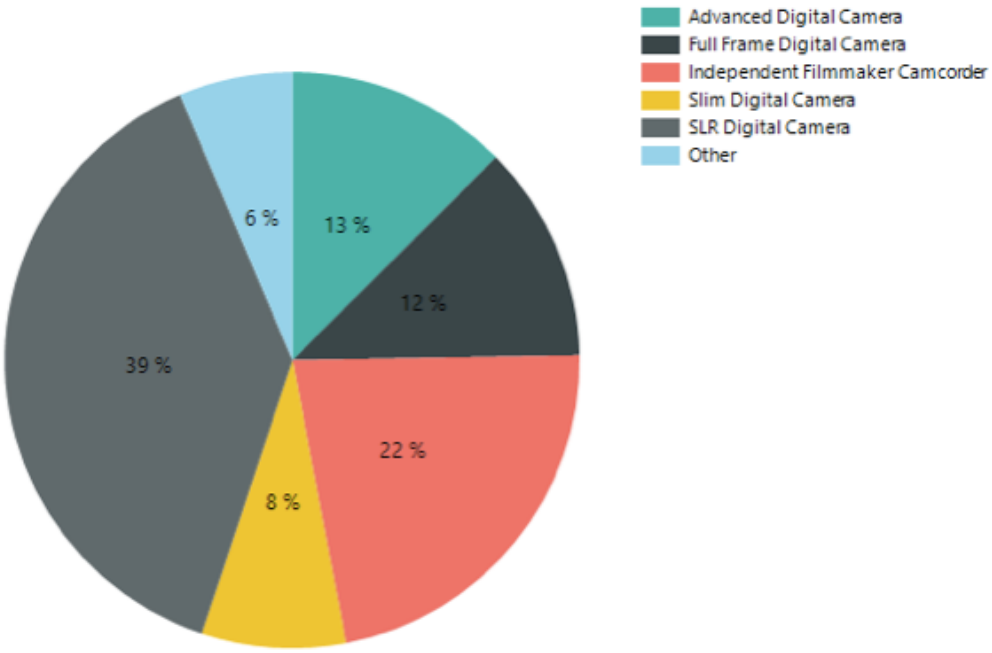

Task Name
Q1 2009
Q2 2009
Q3 2009
Dec '08
Jan '09
Feb '09
Mar '09
Apr '09
May '09
Jun '09
Jul '09
Aug
Planning
Research
Design
Implementation
Follow up

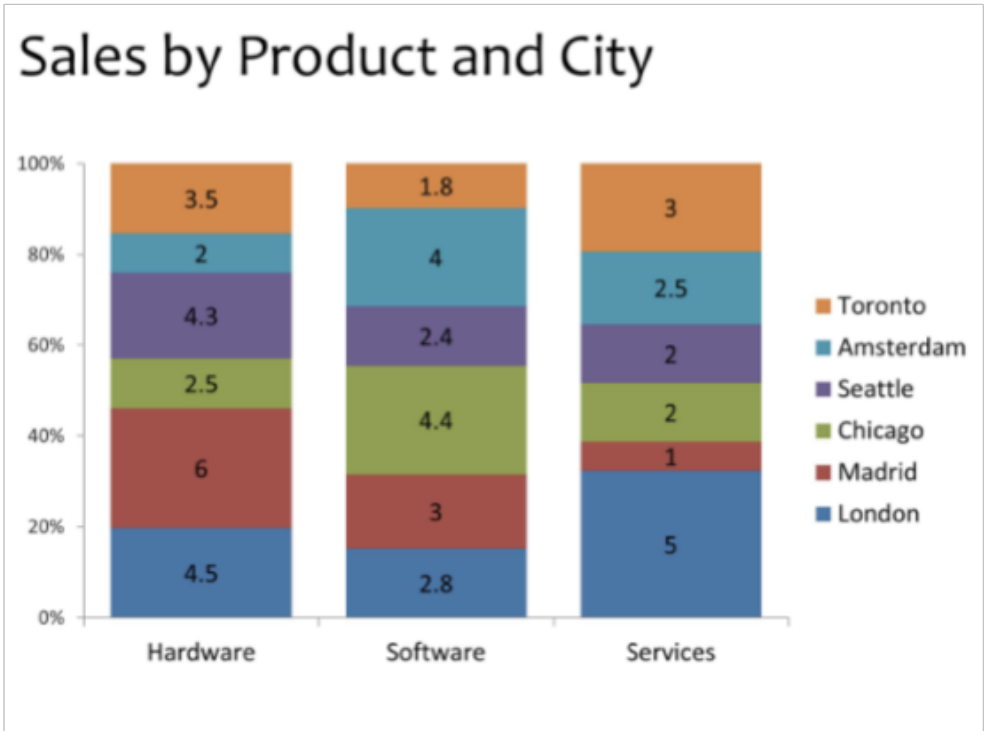
PIE CHART


Who Owns The National Debt?
Total Federal Debt, January 2007: $8,708 B
US Citizens 30.7%
Fed Reserve* 9.0%
Foreigners 25.7%
US Govt. 34.5%
Japanese 7.5%
Chinese 4.1%
Brits 2.9%
OPEC 1.2%
Koreans, Carib Banks, Taiwanese, Germans, Hong Kong, Brazilians, Canadians, Lux, Mexicans, French, Singapore, Swiss, Turks, Irish, Swedes, Dutch, Israelis, Thais, Belgians, Indians, Italians, Poles. Combined= 8.5%
All Others 1.7%
* Estimated breakout of Fed holds vs. US Govt (total $374 B)

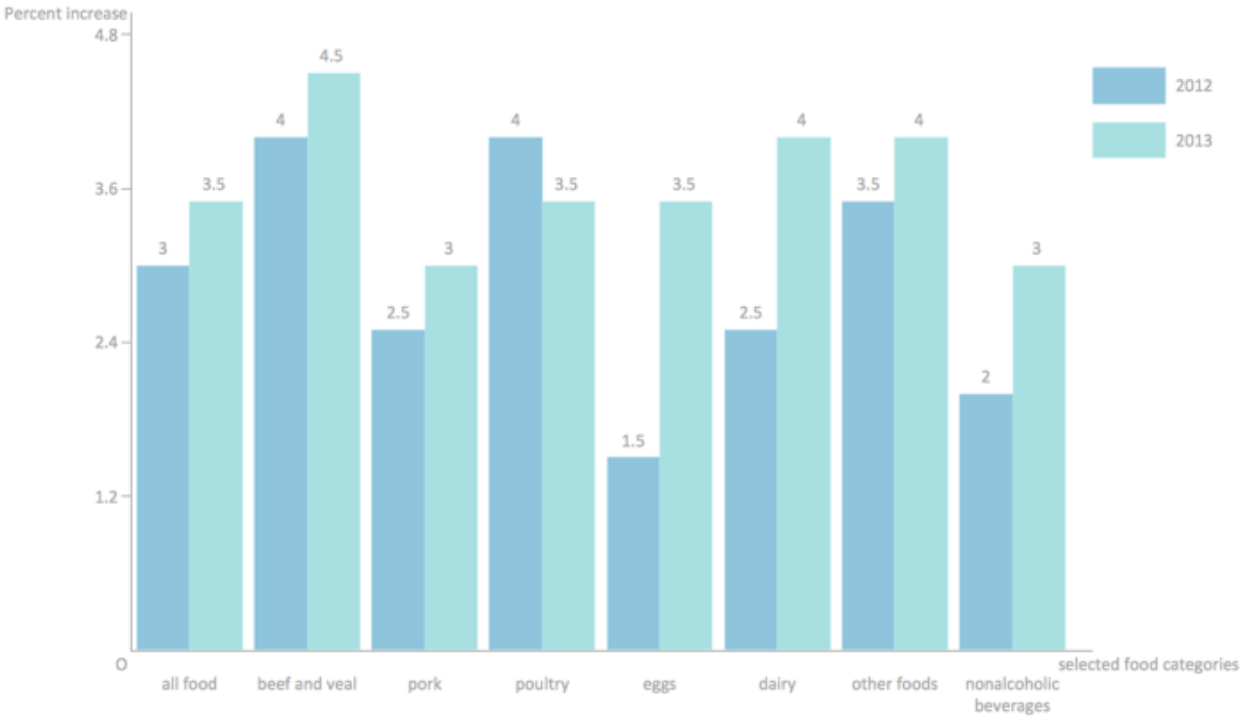
Camera and Camcorder Sales

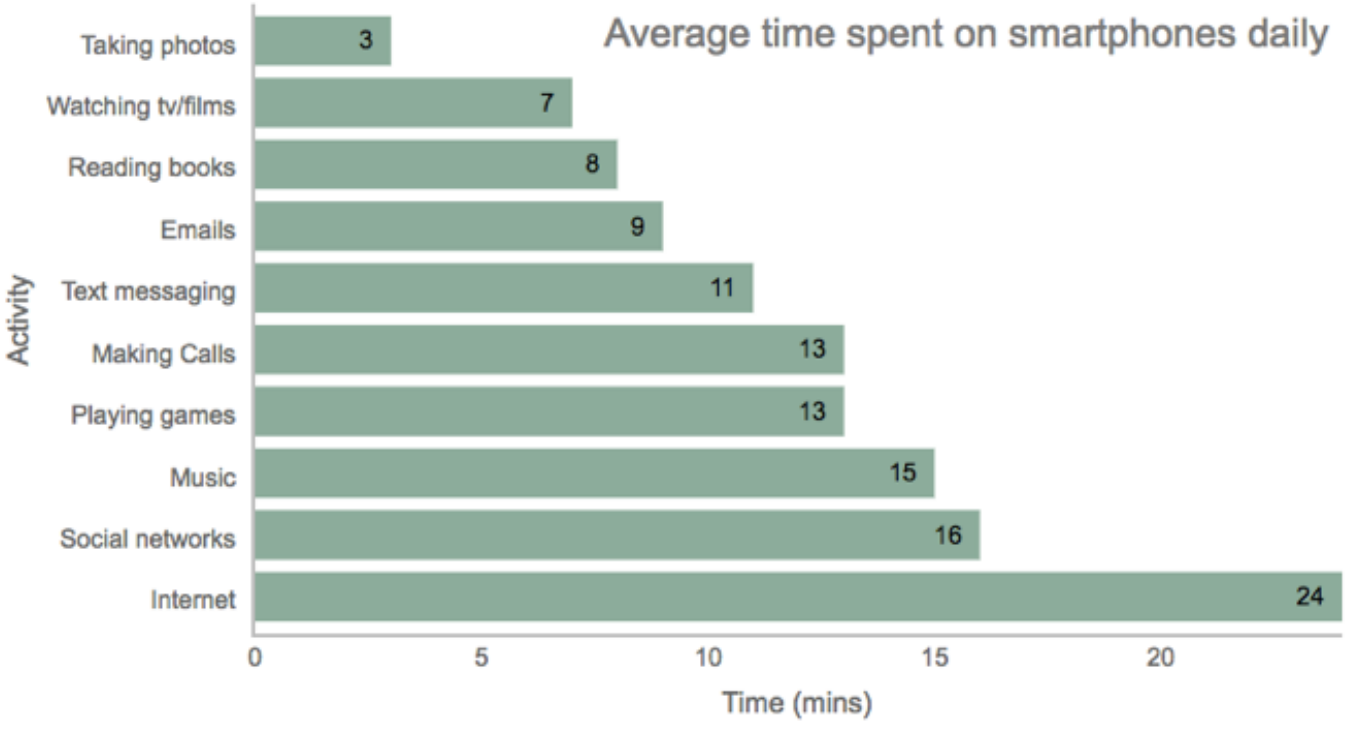
As a Percentage of Total Sales

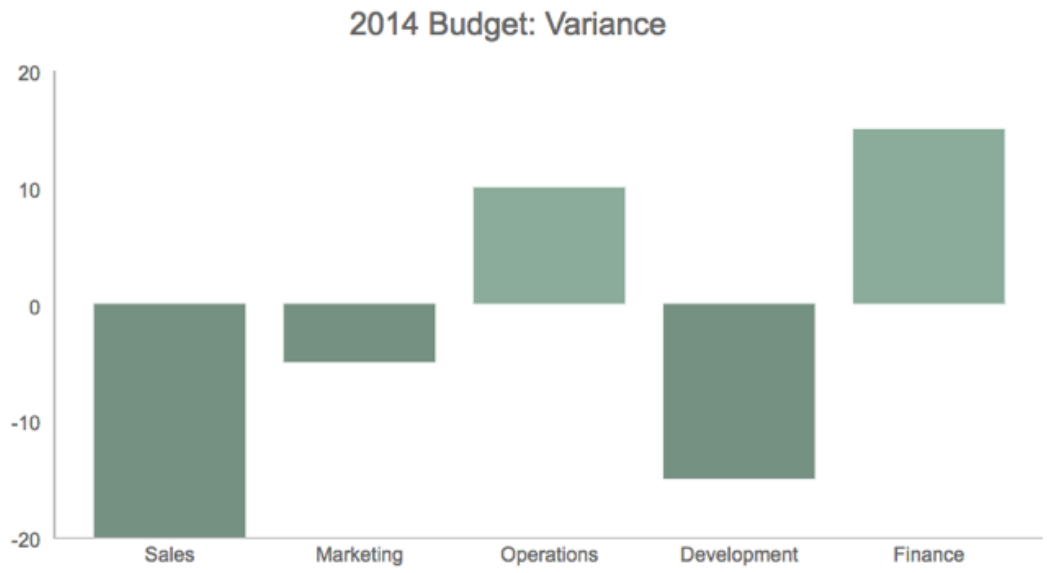




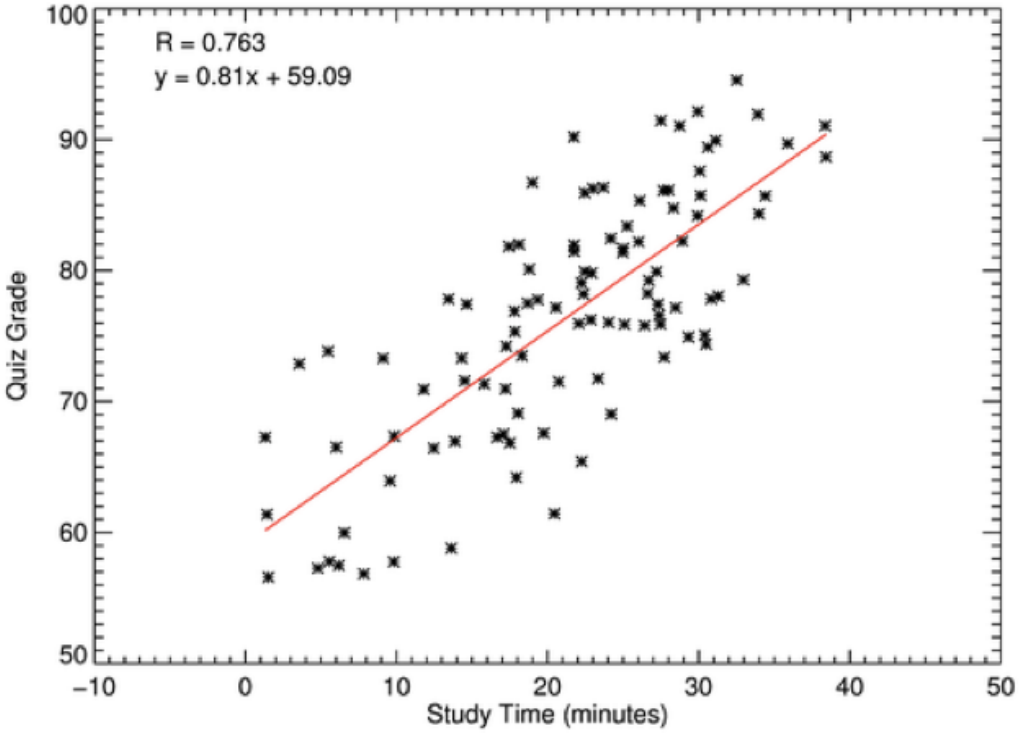
BAR CHART


Percent increase
4.8
3.6
2.4
1.2
0
2012
2013
3
3.5
4
4.5
2.5
3
4
3.5
1.5
3.5
2.5
4
3.5
4
2
3
all food
beef and veal
pork
poultry
eggs
dairy
other foods
nonalcoholic beverages
selected food categories

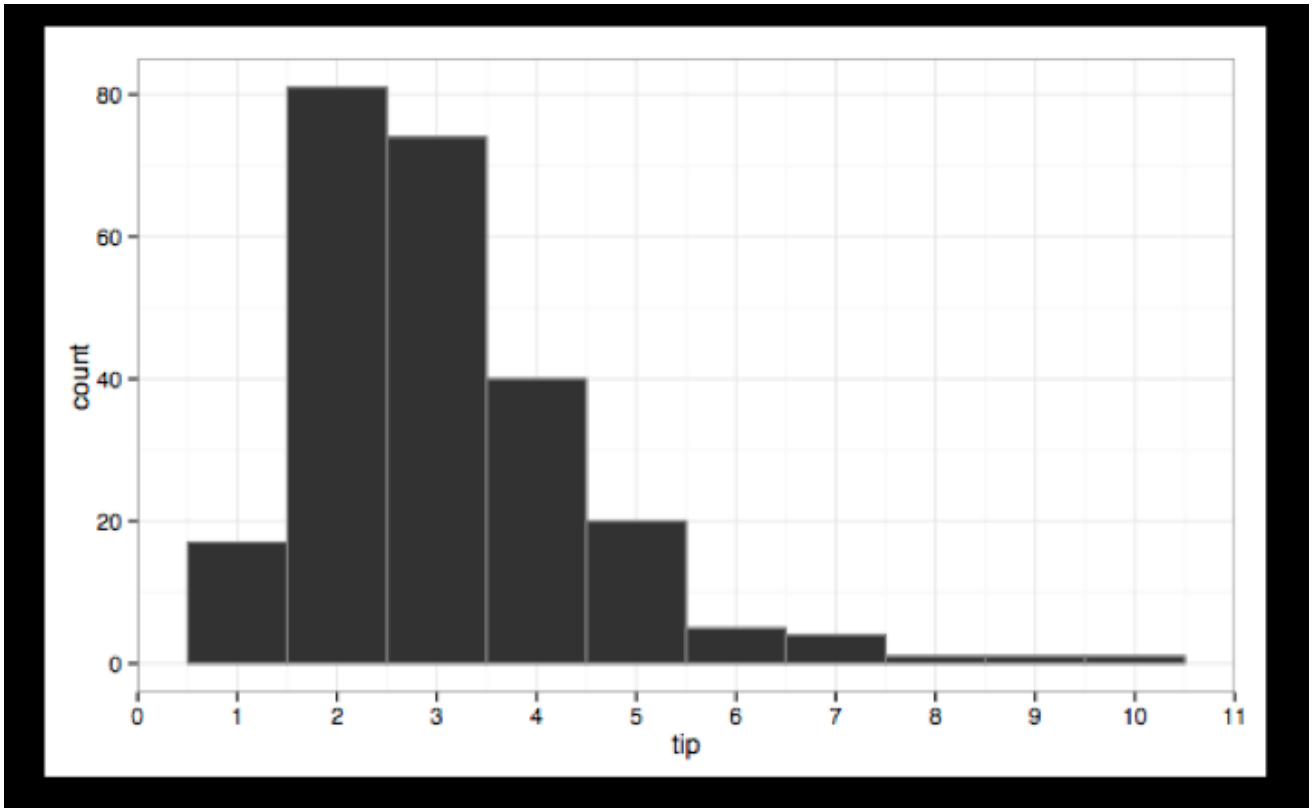

Average time spent on smartphones daily
Activity
Taking photos
3
Watching tv/films
7
Reading books
8
Emails
9
Text messaging
11
Making Calls
13
Playing games
13
Music
15
Social networks
16
Internet
24
0
5
10
15
20
Time (mins)

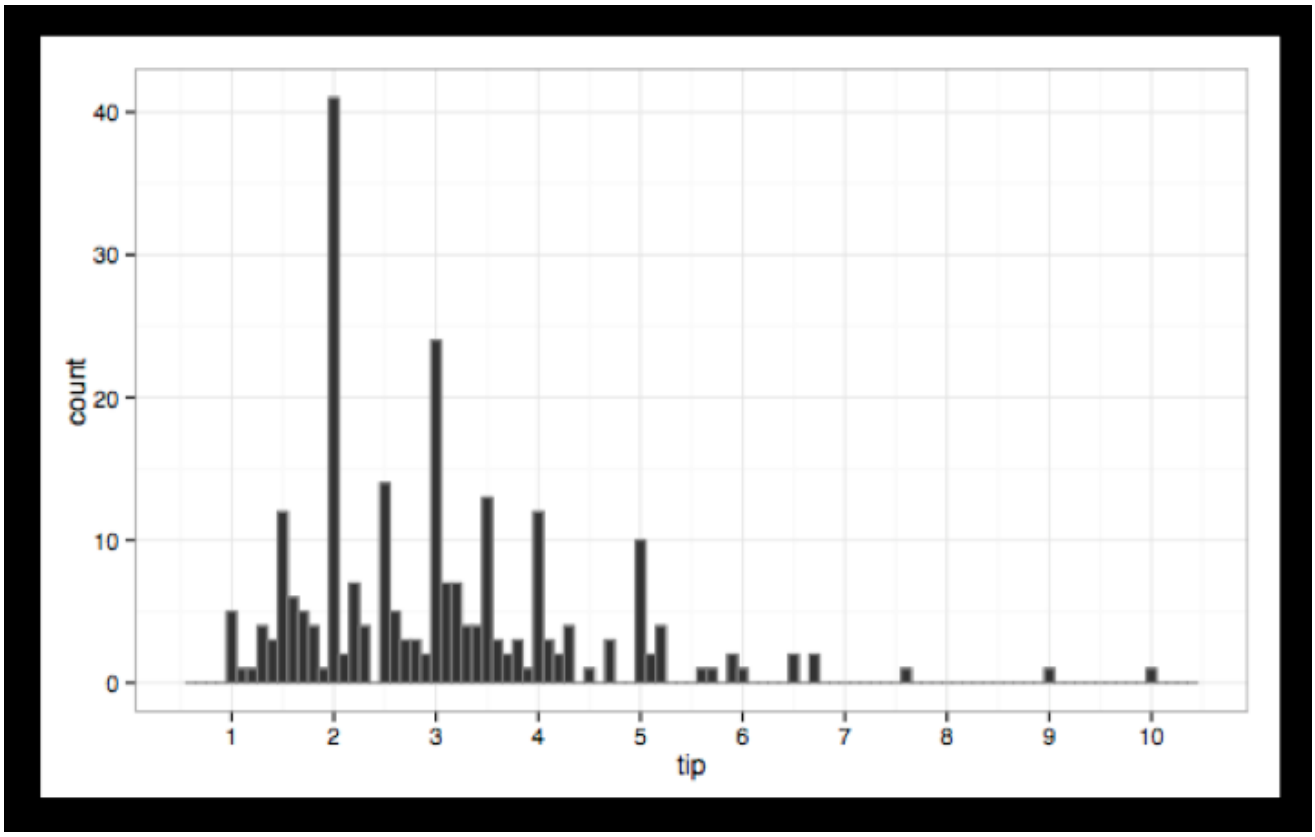


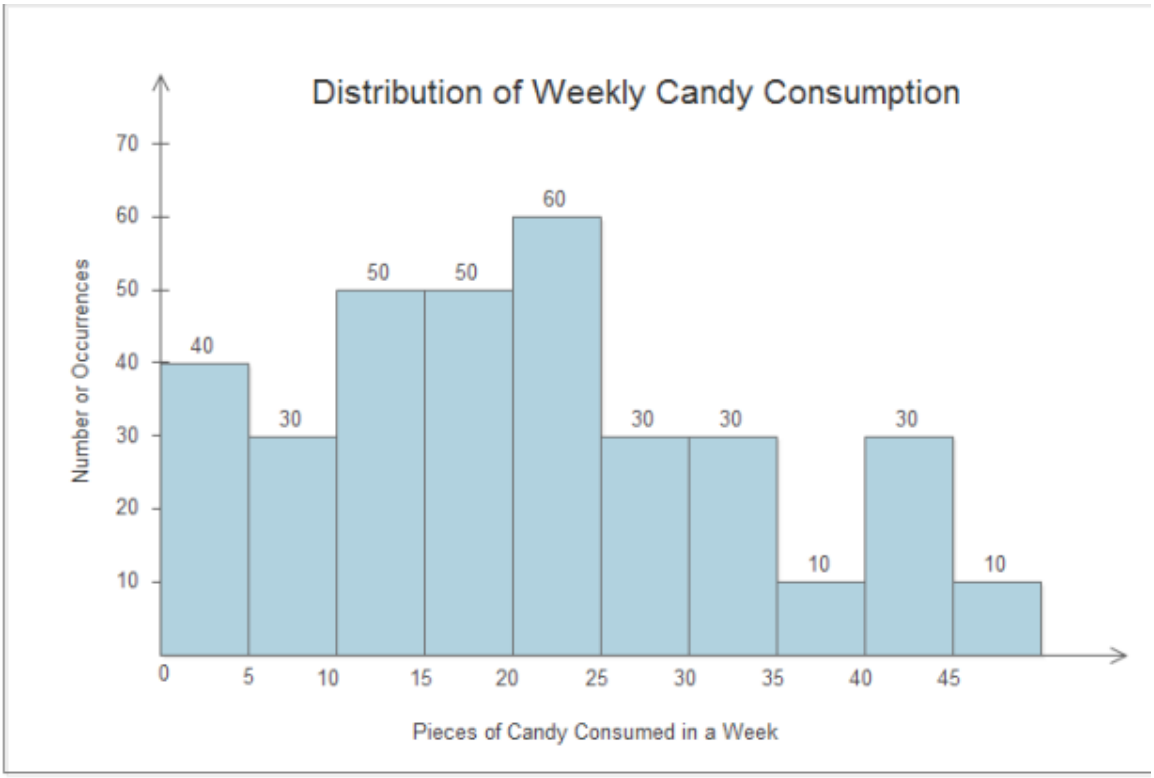
SCATTER PLOT

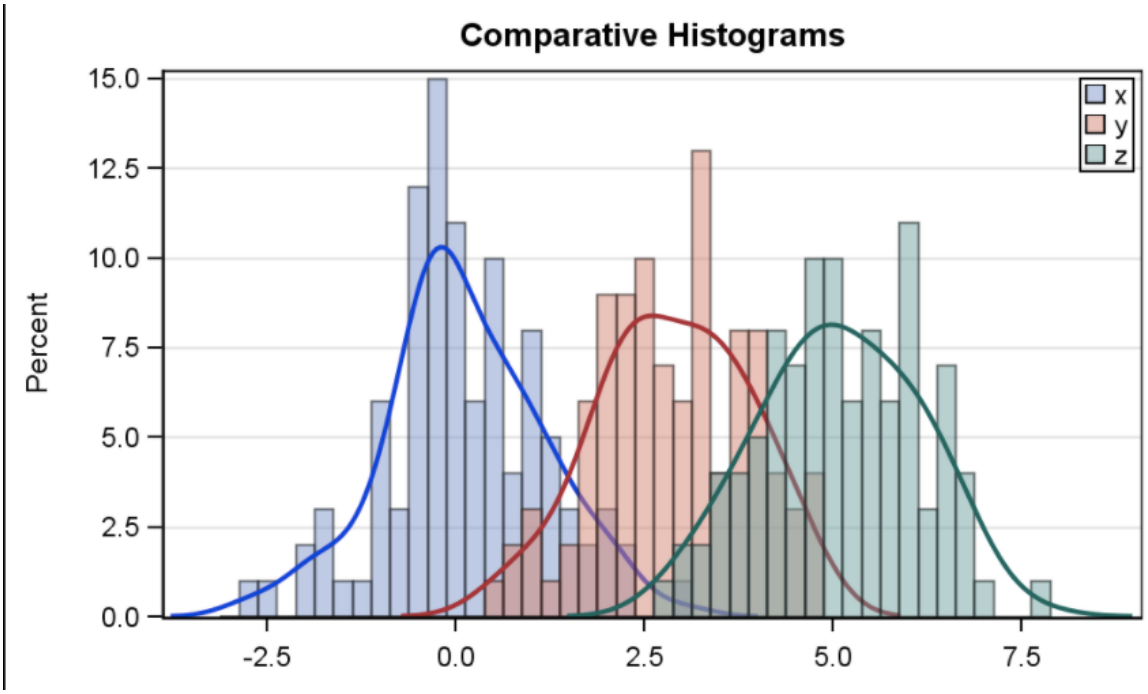


HISTOGRAM

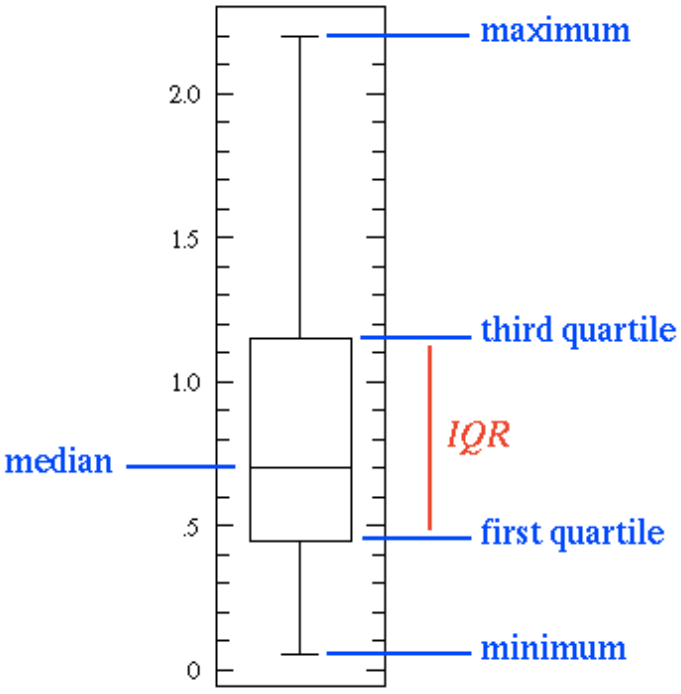

80
60
40
20
0
count
0
1
2
3
4
5
6
7
8
9
10
11
tip

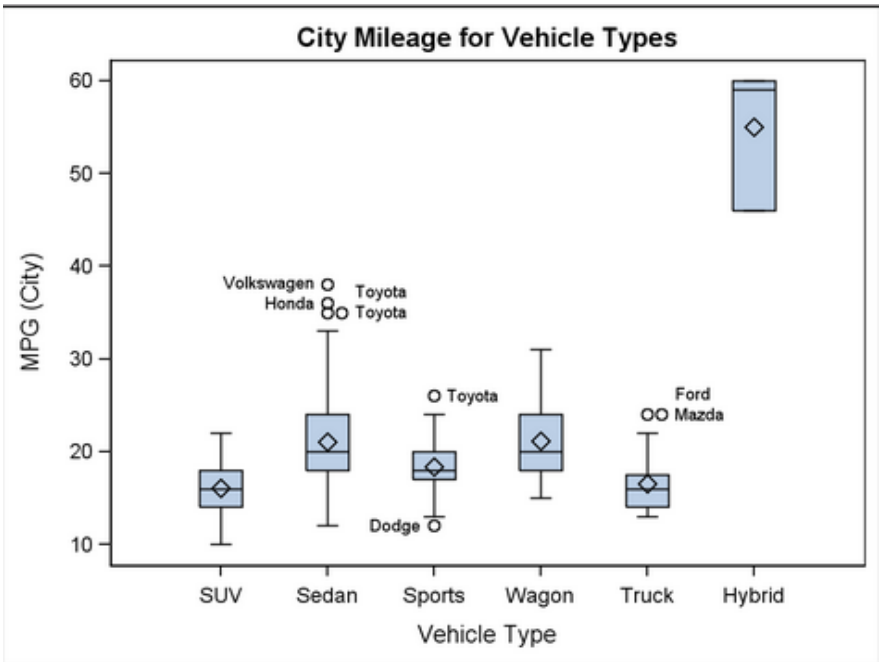

40
30
20
10
0
count
1
2
3
4
5
6
7
8
9
10
tip

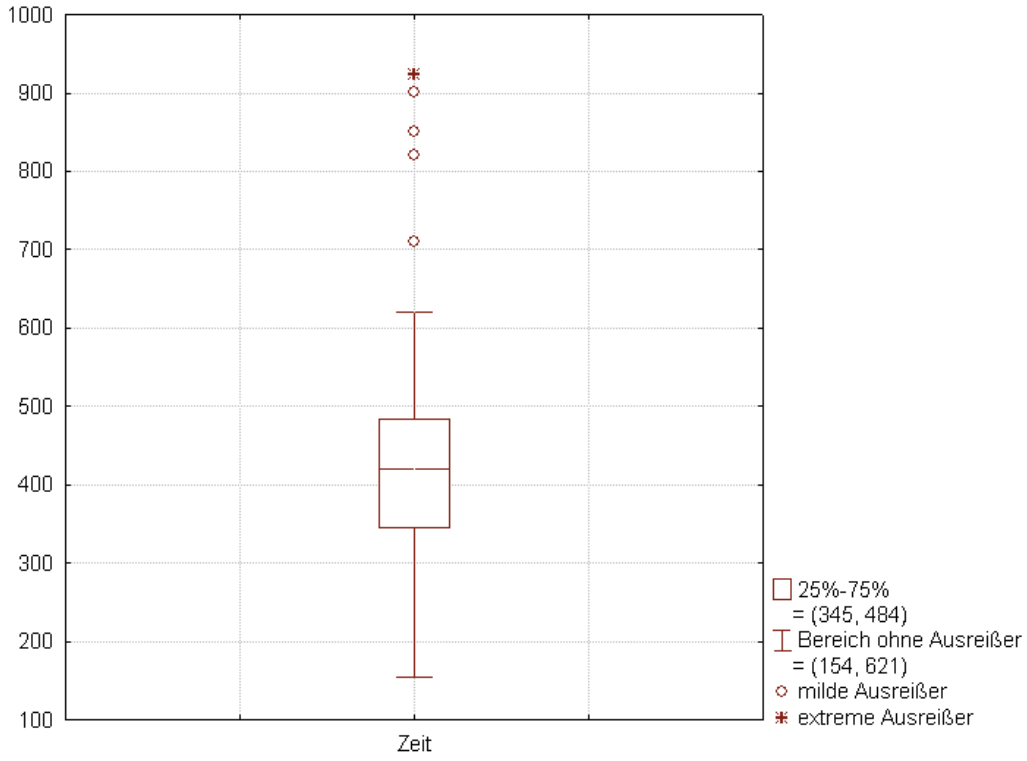

Distribution of Weekly Candy Consumption
Number or Occurrences
70
60
50
40
30
20
10
40
30
50
50
60
30
30
10
30
10
0
5
10
15
20
25
30
35
40
45
Pieces of Candy Consumed in a Week

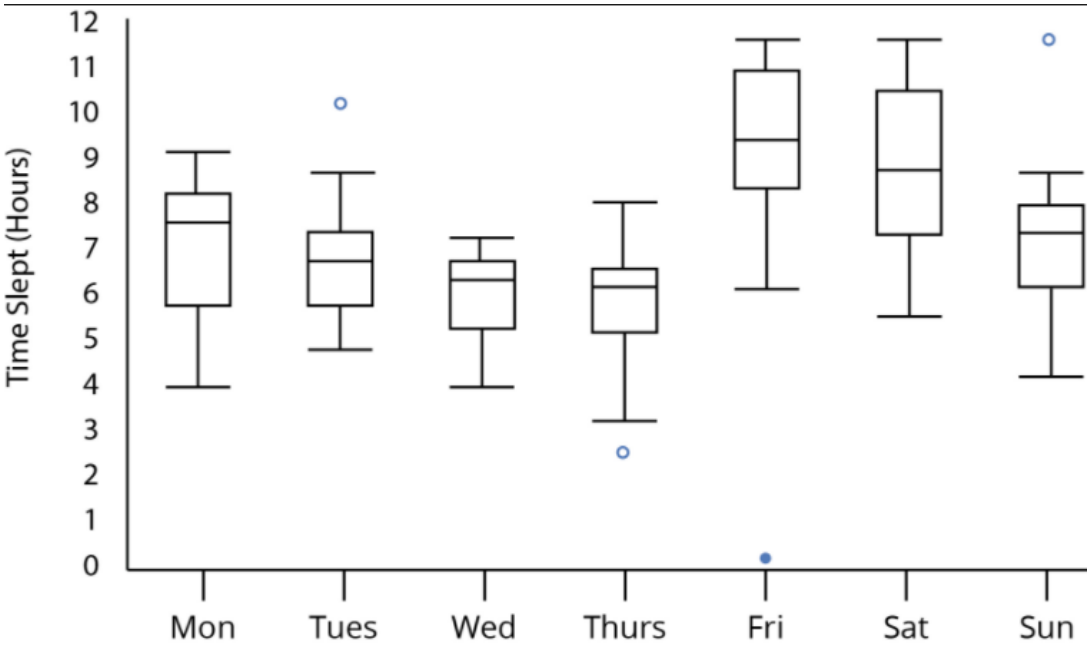

Comparative Histograms
Percent
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-2.5
0.0
2.5
5.0
7.5
x
y
z

BOX AND WHISKER PLOT

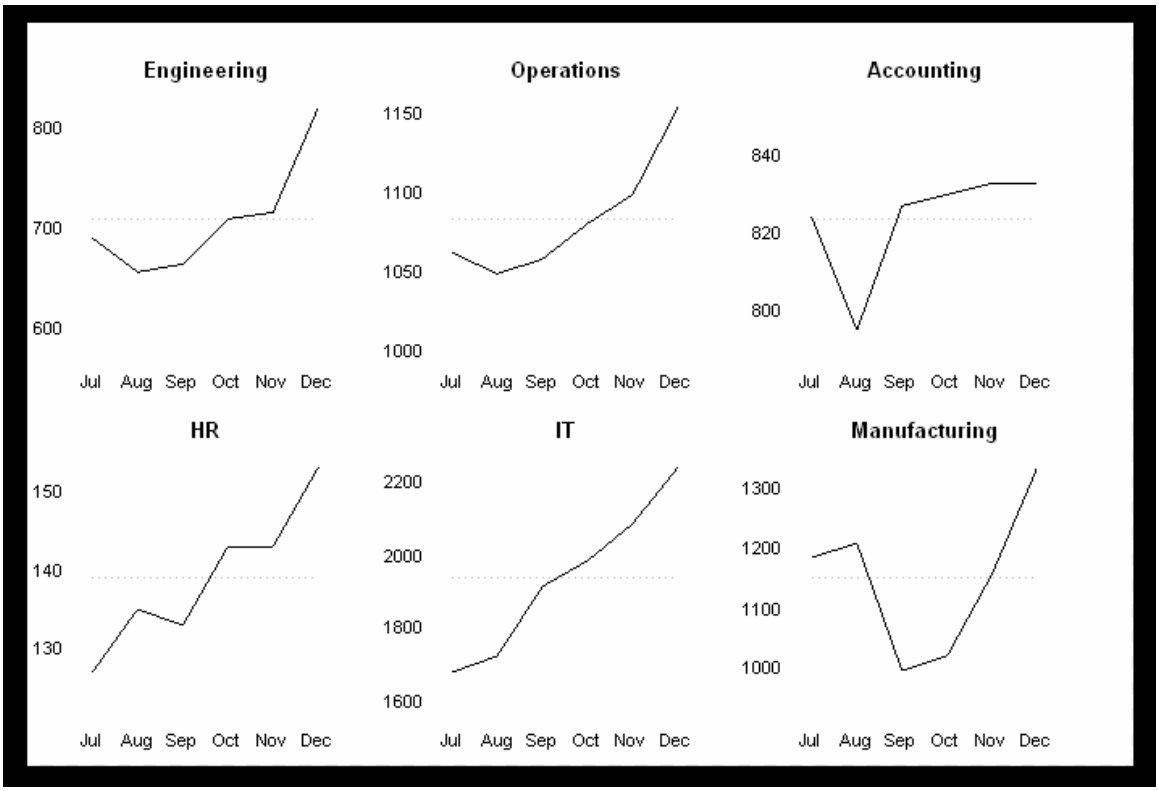




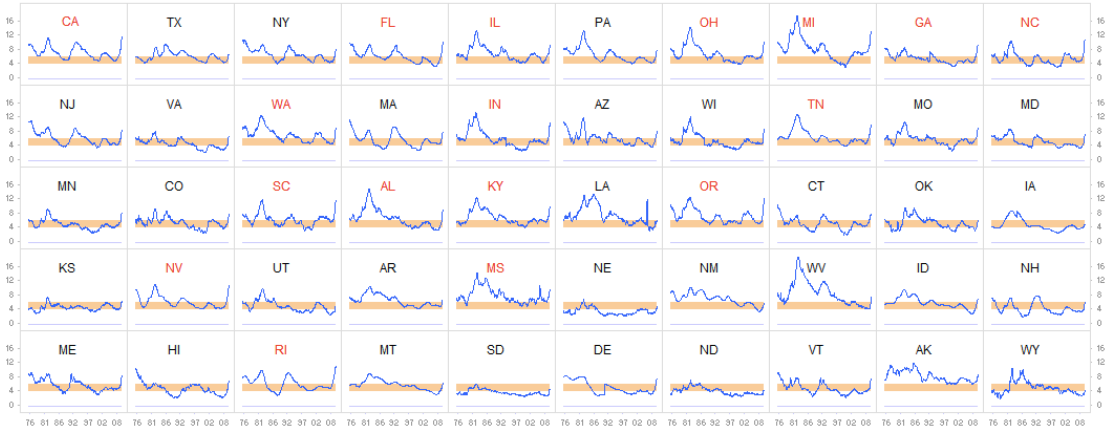

1000
900
800
700
600
500
400
300
200
100
Zeit
25%-75%
= (345, 484)
Bereich ohne Ausreißer
= (154, 621)
milde Ausreißer
extreme Ausreißer


Time Slept (Hours)
12
11
10
9
8
7
6
5
4
3
2
1
0
Mon
Tues
Wed
Thurs
Fri
Sat
Sun

SMALL MULTIPLES



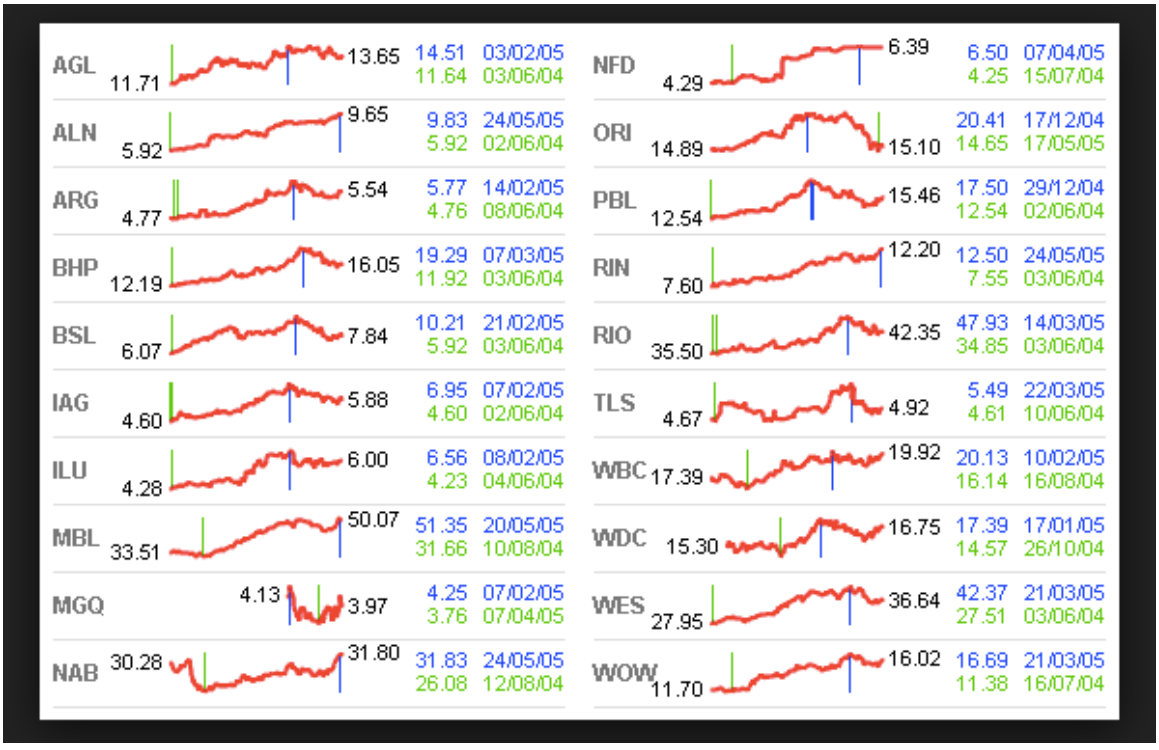
Monthly Unemployment Rates by State, Jan 1976 - Apr 2009

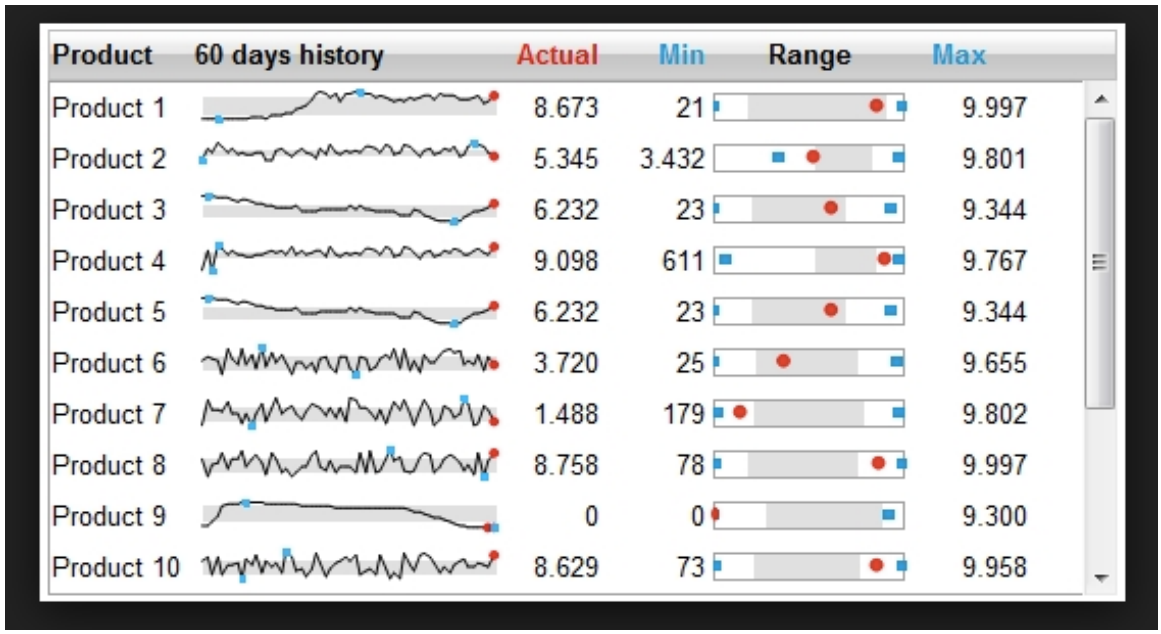


Source: Bureau of Labor Statistics

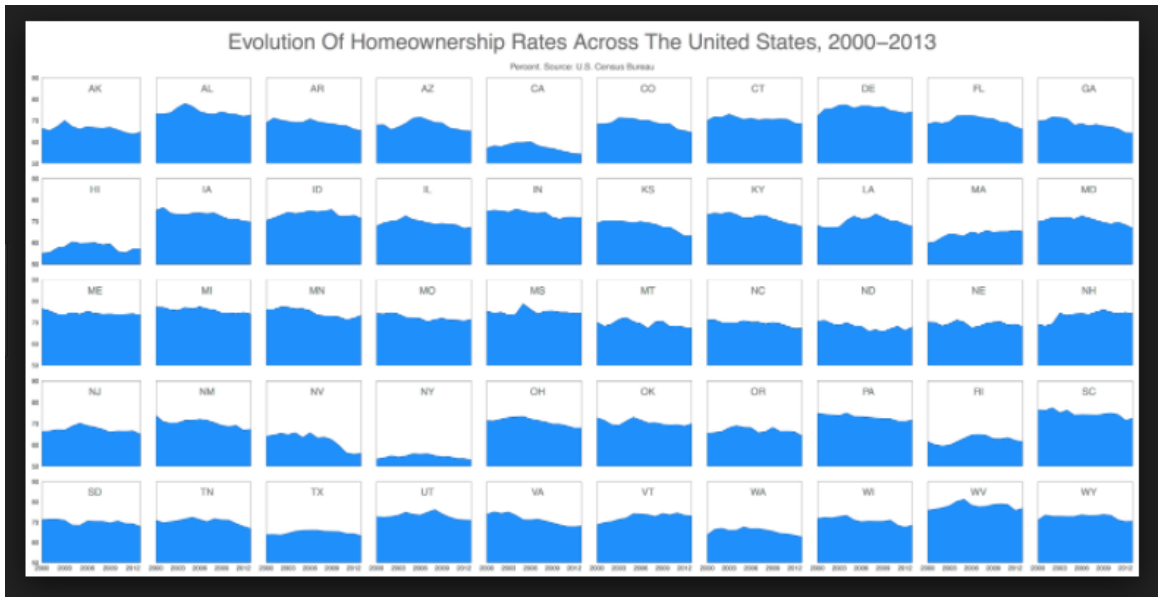
Notes: The orange band denotes a "normal" unemployment rate (4%-6%);
State code in red: unemployment rate in April 2009 is higher than the US average


2000: State-level support (orange) or opposition (green) on school vouchers, relative to the national average of 45% support
Income under $20,000
$20-40,000
$40-75,000
$75-150,000
Over $150,000
All voters
White Catholics
White evangelicals
White non-evang. Protestants
White other/ no religion
Blacks
Hispanics
Other races
Orange and green colors correspond to states where support for vouchers was greater or less than the national average.
The seven ethnicreligious cagetories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants.
Where a category represents less than 1% of the voters of a state, the state is left blank.


AGL 11.71 13.65 14.51 03/02/05 11.64 03/06/04
ALN 5.92 9.65 9.83 24/05/05 5.92 02/06/04
ARG 4.77 5.54 5.77 14/02/05 4.76 08/06/04
BHP 12.19 16.05 19.29 07/03/05 11.92 03/06/04
BSL 6.07 7.84 10.21 21/02/05 5.92 03/06/04
IAG 4.60 5.88 6.95 07/02/05 4.60 02/06/04
ILU 4.28 6.00 6.56 08/02/05 4.23 04/06/04
MBL 33.51 50.07 51.35 20/05/05 31.66 10/08/04
MGQ 4.13 3.97 4.25 07/02/05 3.76 07/04/05
NAB 30.28 31.80 31.83 24/05/05 26.08 12/08/04
NFD 4.29 6.39 6.50 07/04/05 4.25 15/07/04
ORI 14.89 15.10 20.41 17/12/04 14.65 17/05/05
PBL 12.54 15.46 17.50 29/12/04 12.54 02/06/04
RIN 7.60 12.20 12.50 24/05/05 7.55 03/06/04
RIO 35.50 42.35 47.93 14/03/05 34.85 03/06/04
TLS 4.67 4.92 5.49 22/03/05 4.61 10/06/04
WBC 17.39 19.92 20.13 10/02/05 16.14 16/08/04
WDC 15.30 16.75 17.39 17/01/05 14.57 26/10/04
WES 27.95 36.64 42.37 21/03/05 27.51 03/06/04
WOW 11.70 16.02 16.69 21/03/05 11.38 16/07/04



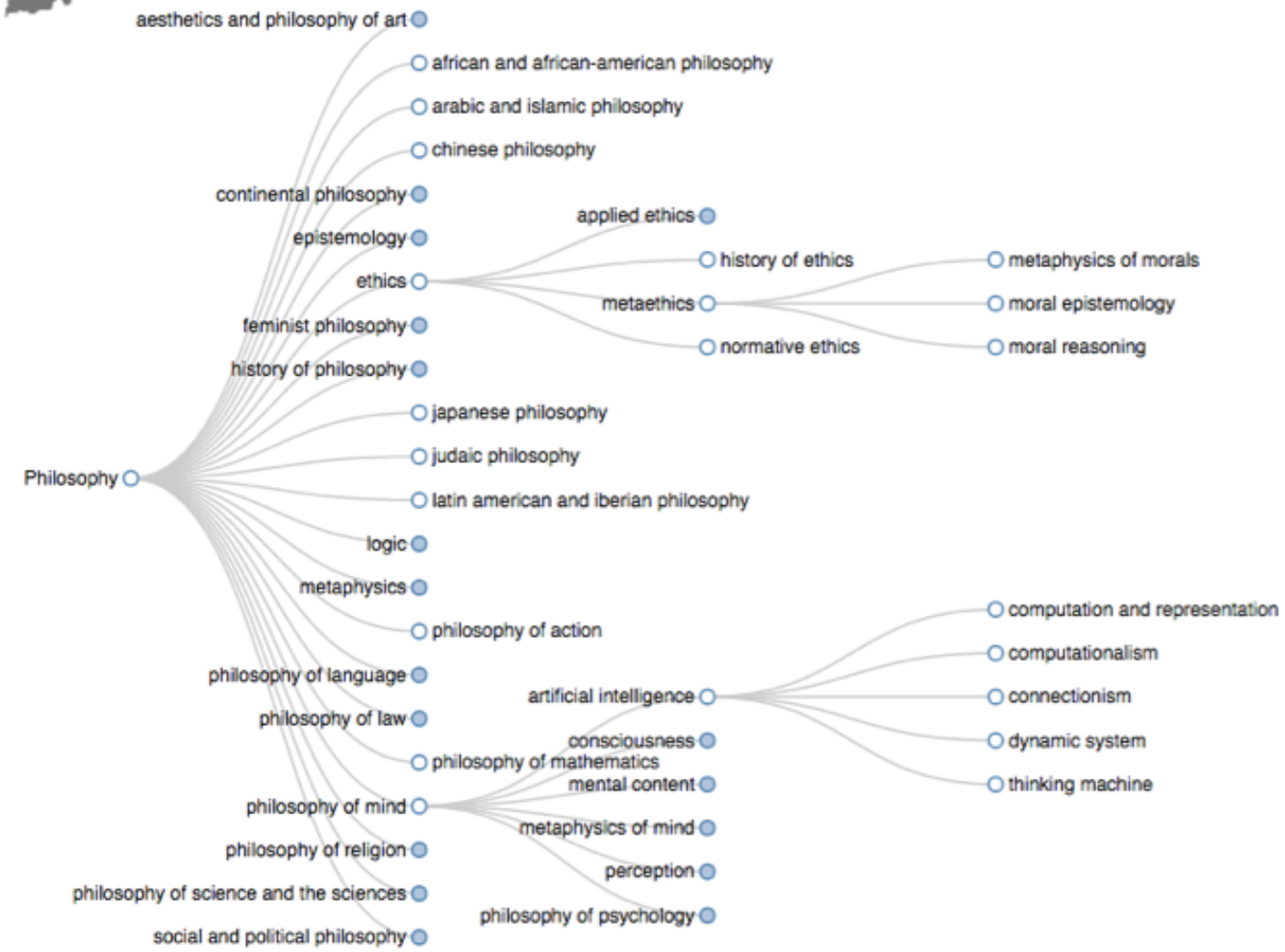
Since small multiples are visually dense, the use of area in this graph makes noticing small differences in the graph more difficult.

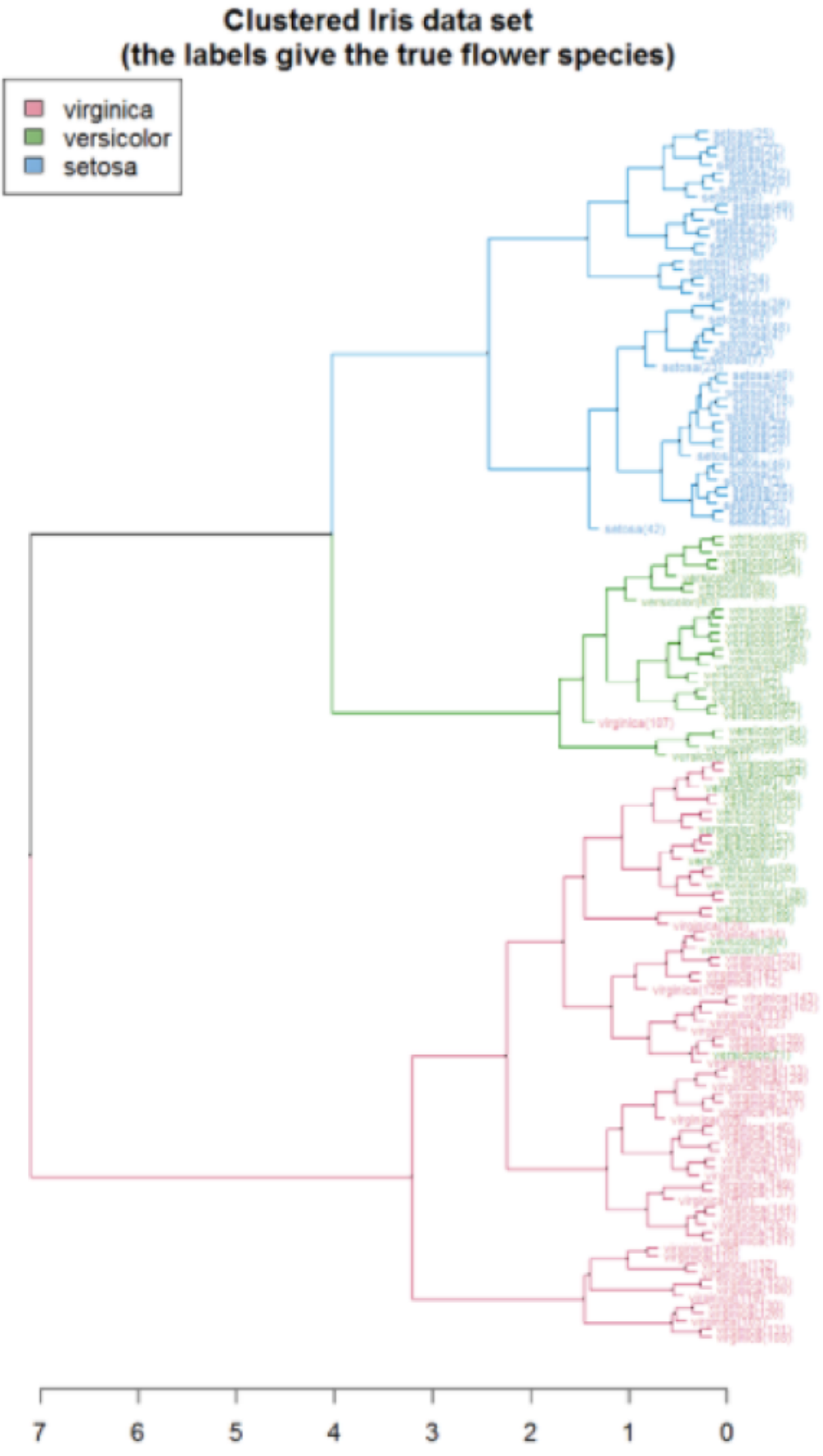


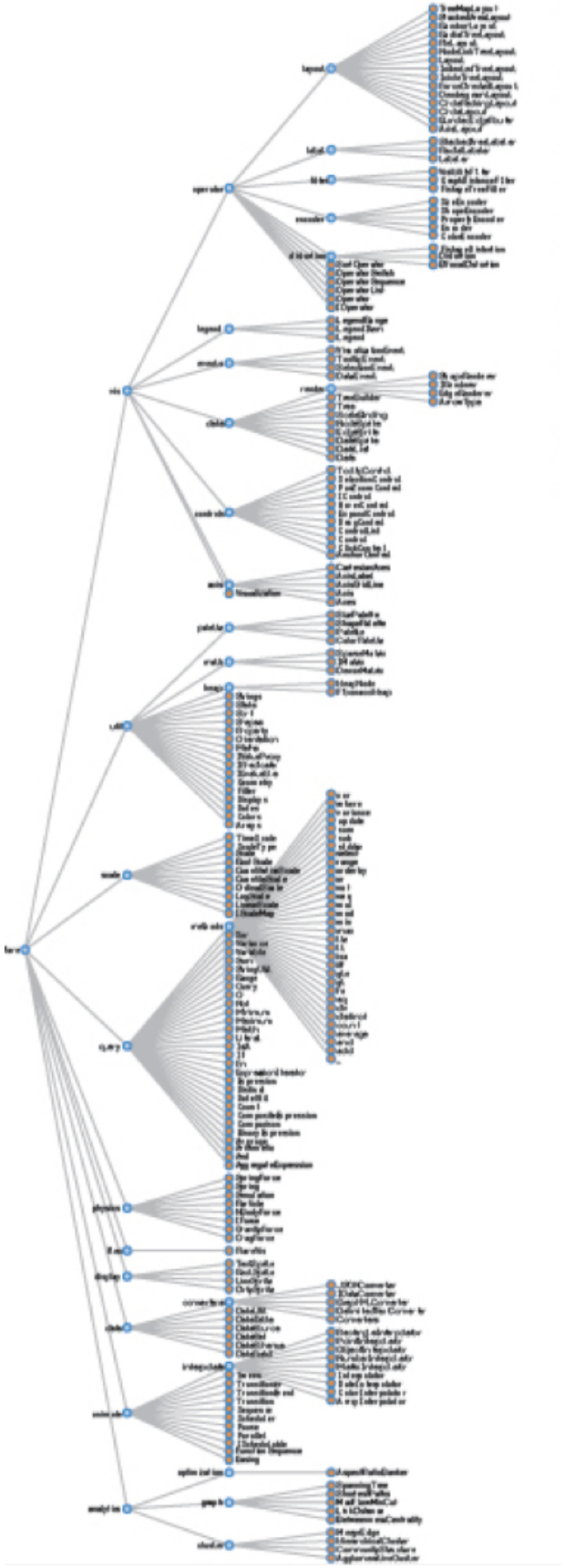
HIERARCHICAL TREE


```
/local $  source ~/Development/virtual_environments/dauer/bin/activate
```

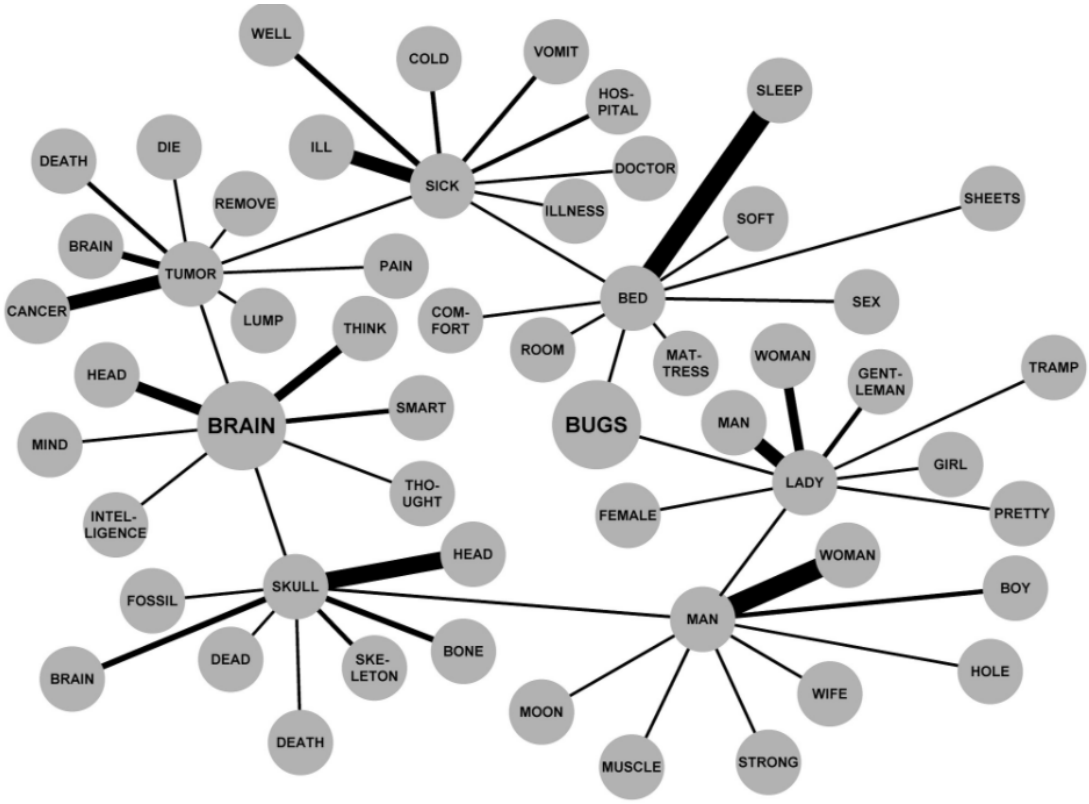
InPhO : The Taxonomy

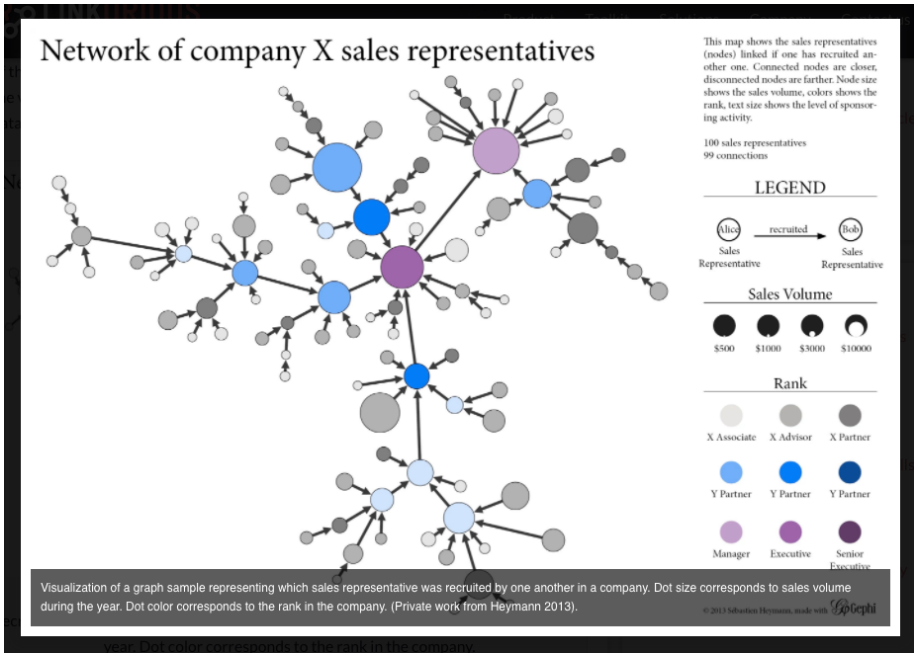



Clustered Iris data set
(the labels give the true flower species)
virginica
versicolor
setosa
7
6
5
4
3
2
1
0



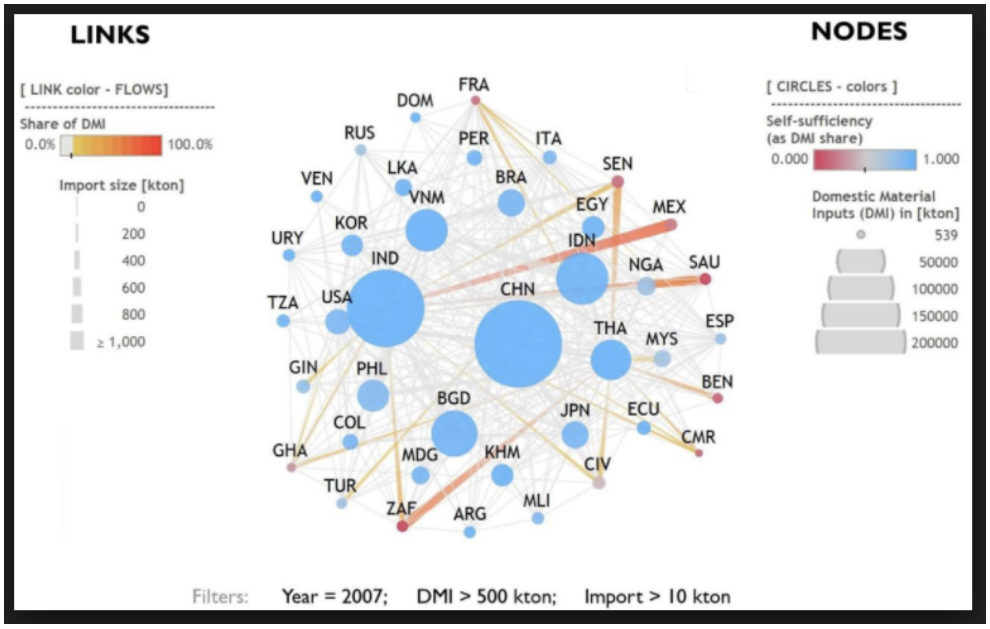
NODE LINK DIAGRAM



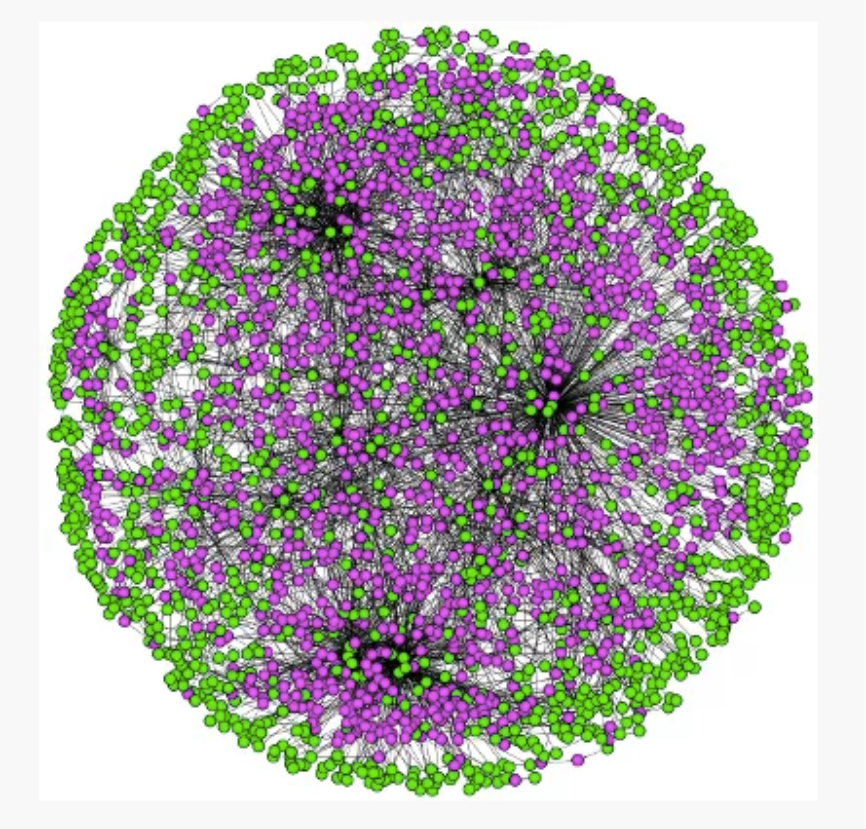


Visualization of a graph sample representing which sales representative was recruited by one another in a company. Dot size corresponds to sales volume during the year. Dot color corresponds to the rank in the company. (Private work from Heymann 2013).

A node link diagram with line color/thickness representing an additional variable.



The Dreaded Hairball!!



TAXONOMY OF VISUALIZATION METHODS

One-Dimensional

- List (see spreadsheet)

Geospatial (Two-Dimensional)

- Dot Distribution (see spreadsheet)
- Cartogram (used for mapping categorical data; a geographic region on the map is distorted to represent the mapping variable's measurement)
- Choropleth (used for mapping continuous data)
- Contour Line (links points of the same value by attribute – e.g., a map with every point of the same elevation displayed using the same color ring)

Three-Dimensional

- Volume Rendering (displaying multiple, consecutive slices of 2D images to create the illusion of three dimensions; used to visualize 3D objects that cannot be seen by the naked eye because of size or barrier)
- Computer Generated Modeling (spatial modeling of a 3D dataset; does not necessarily have to be an object)

Temporal

- Timeline (discrete steps over time)
- Time series (see spreadsheet; continuous variable measured over time)
- Gantt (overlapping steps in a process, over time)
- Stream Graph (displays part to whole and total measurements over time; used in place of stacked bars)

- Sankey Diagram (flow diagram in which the width of arrows is proportional to their flow quantity/intensity)
- Alluvial Diagram (flow diagram documenting the change in network structures over time)

Multi-Dimensional

- Pie Chart (see spreadsheet)
- Bar Chart (see spreadsheet)
- Scatter Plot (see spreadsheet)
- Histogram (see spreadsheet)
- Tree Map (displays hierarchical data using rectangles whose sizes are proportional to the value of the variable being measured; rectangles are nested to denote set/subset relationships)
- Step Chart (time series; most effectively displays segmented data – e.g., stamp prices only change every few years)
- Area Chart
- Heat Map (individual values in a matrix are represented as colors)
- Radar/Spider Chart
- Box plot (see spreadsheet)
- Parallel Coordinates (use for nominal data values across multiple trials/ measurements)
- Waterfall Chart
- Small Multiples

Tree/Hierarchical

- Basic Tree Diagram
- Dendogram
- Radial Tree
- Hyperbolic Tree (use when there are many terminal nodes)
- Tree Map

Network

- Node-Link Diagram
- Dependency Graph/Circular Diagram
- Tube Map