Visual Analytics

*Misleading statistical graphs*

Find from the web, at least 10 plots for which you think that something went wrong.  Explain what is this and make if possible the correct plot. In case the data are not available you may just describe and sketch what should be the plot.

Professor: D. Karlis

Angeliki Chatzimoschou (BAPTXXXX)

2/2/2017

# Contents

[3D Effects 2](#_Toc474165324)

[Biased results 2](#_Toc474165325)

[AND Pie chart AND overloaded 3](#_Toc474165326)

[Bollywood 4](#_Toc474165327)

[Y- axis problem 4](#_Toc474165328)

[Full reality distortion 5](#_Toc474165330)

[Inaccurate lengths 6](#_Toc474165331)

[Stacked columns 7](#_Toc474165332)

[Totally irrelevant chart 7](#_Toc474165333)

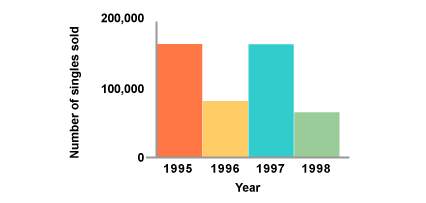
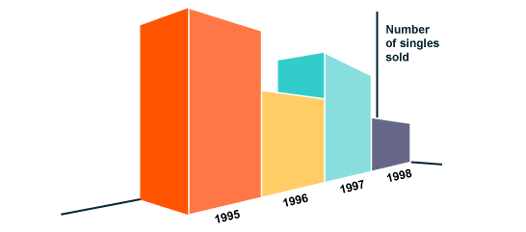
[Wrong scaling 8](#_Toc474165334)

[Pictographs 8](#_Toc474165335)

Visual Analytics

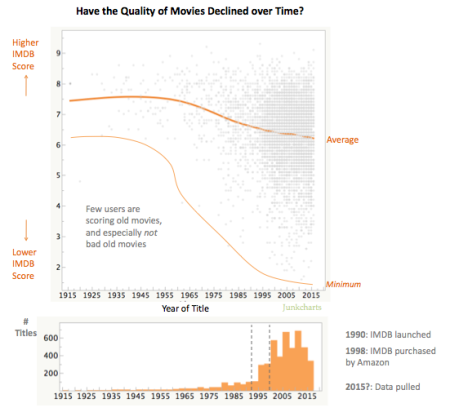
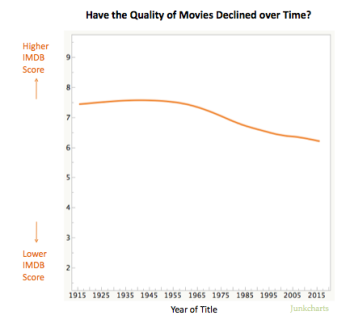
Misleading statistical graphs

# **3D Effects**



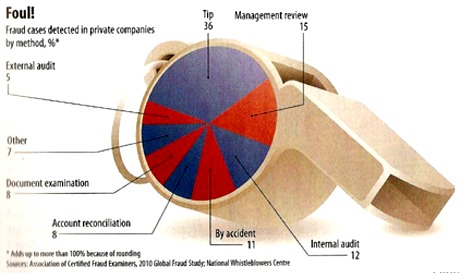
In the 3D plot the scale of the Y-axis is missing, thus it gives the impression that sales in 1995 are more than all the rest of the years, while actually when representing the data in a 2D barplot, we clearly observe that years: 1995 and 1997 share the same amount of sales.

# **Biased results**



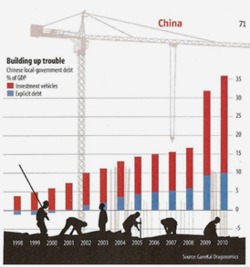
Observing the first line plot, we may easily and wrongly conclude that since the average film scored from 7.5 to 6.5 the quality of movies as reflected by IMDB, has declined over time. This doesn’t mean that IMDB users like old movies more than the recent ones, since IMDB's website launched in 1990. This effects the users’ reviewing movie list, as it is more likely to rate a film you recently watched than an older one. So, we have a bias through time issue to address on the first plot, which is explained better on the second plot.

# **AND Pie chart AND overloaded**

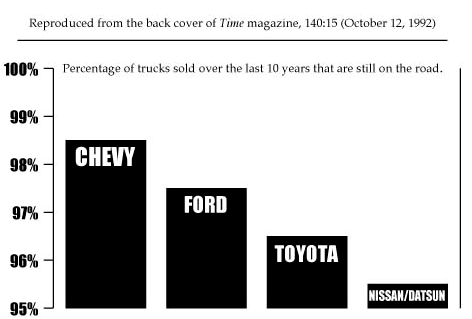


In general, it is not a good idea to represent data using pie charts as it makes it difficult to apprehend fast and clear, the proportional difference among the slices.Here, apart from the distorted pie slices, we also have to deal with extra unnecessary decorations that keep the eye distracted from the essential information. A simple 2D barplot would be a much better option.

# **Bollywood**

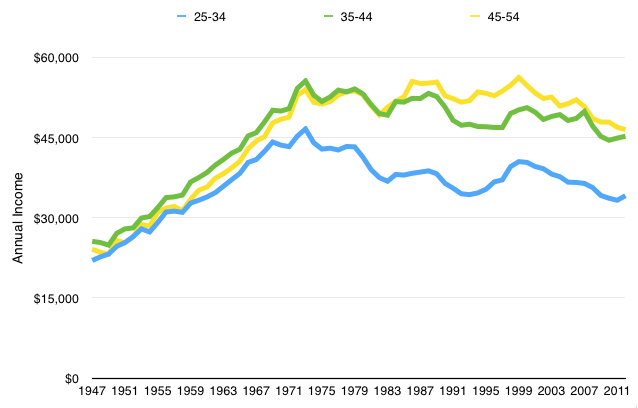
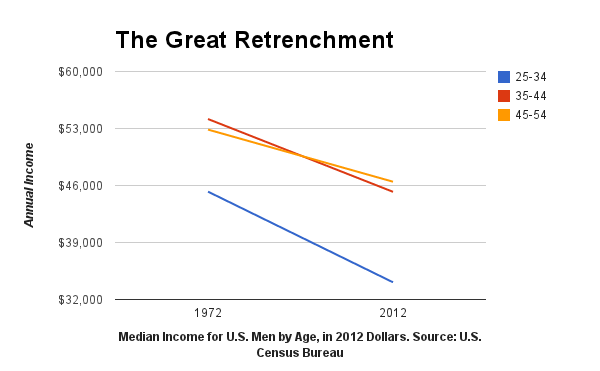
  
  
The crane and workers are redundant as they offer nothing to the meaning of the chart and make it more difficult to focus on the actual barplot. By isolating the stacked barplot we get a much better idea of what’s going on.

# **Y- axis problem**

* 
* This bar chart indicates that Chevy sales are double than Toyota’s and way larger than Nissan/Datsun’s. This is a misleading graph as the Y-axis doesn’t begin from 0 and gives a totally distorted impression over the sales of each brand. The proper barplot should have a Y-axis beginning from 0 and offer a more balanced impression as the difference between 95% and 98.5% isn’t as huge as it seems here.

# **No color levels** http://i2.wp.com/flowingdata.com/wp-content/uploads/2012/11/Divided-nation.jpg?fit=600%2C448 This map plot was published in a newspaper and obviously, there is a couple of problems. The corrected map should have different colors for each candidate separately and also degrading colors to depict the voting level for each state.

# **Full reality distortion**

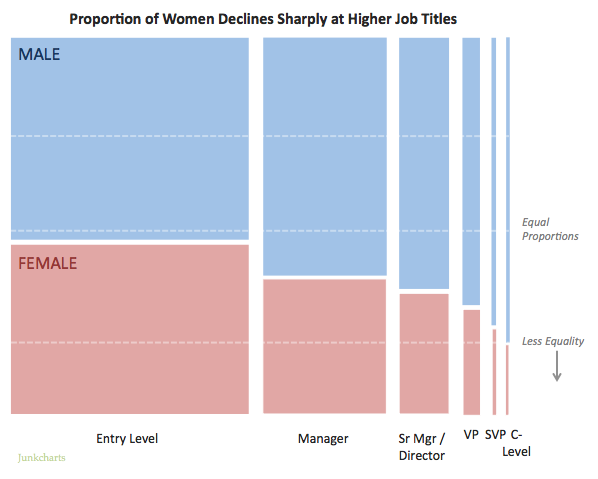
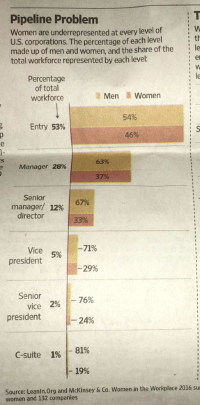
  
  
In a nutshell, in the first line graph Bloomberg is trying to support its notion that during the past 30 years (from 1972 to 2012) the annual median income for US men is following a severely downward trend. It uses the U.S. Census dataset. There are several reasons that this plot is seriously wrong and distorts reality. First of all, it uses lines instead of time series and is difficult to accurately pinpoint the income variances through the years. Secondly, the Y-axis doesn’t begin from 0 and thus makes the slope of the lines seem greater than it actually is. Finally, the X-axis contains only 2 data points and we lose all the in-between detailed information.

The second graph which uses all of the dataset, since 1947, corrects all the mistakes made in the first graph and concludes to the fact that the reality is totally different. There’s an overall income growth with more of a slightly decreasing trend from 1972 to 2012.

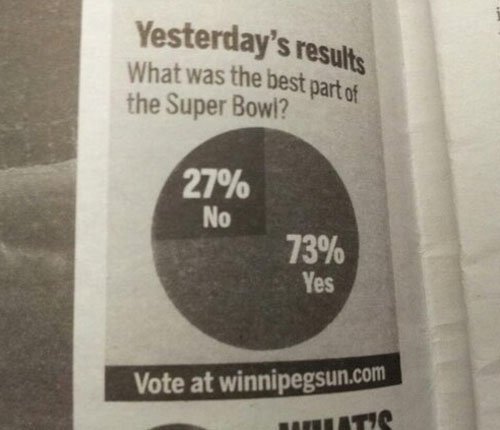
# **Inaccurate lengths** Nbc_olympicmedalshttp://junkcharts.typepad.com/.a/6a00d8341e992c53ef01bb092a5b59970d-pi

In the first plot, although China and Russia have gained the same number of medals, the bar lengths are different. Also, the fact that the names of the countries are in bold, distracts the viewer and focuses more on the opponents than the number of medals won, which is the actual point of the graph. In the second graph, we get a more appropriate and distinct representation.

# **Stacked columns**

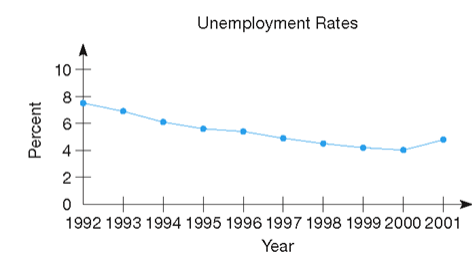
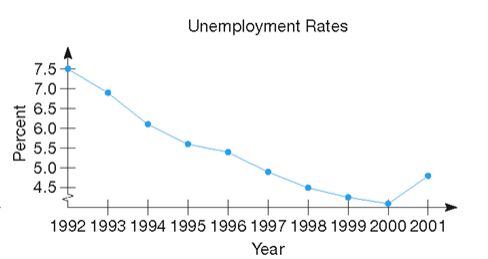
  
  
The creator is trying to use an alternative for the horizontal stacked bar chart but instead it ends up giving stacked columns in each bar. Since there is a double message to pass, the gender inequality and the imbalance between job levels, a mosaic plot is a better choice.

# **Totally irrelevant chart**

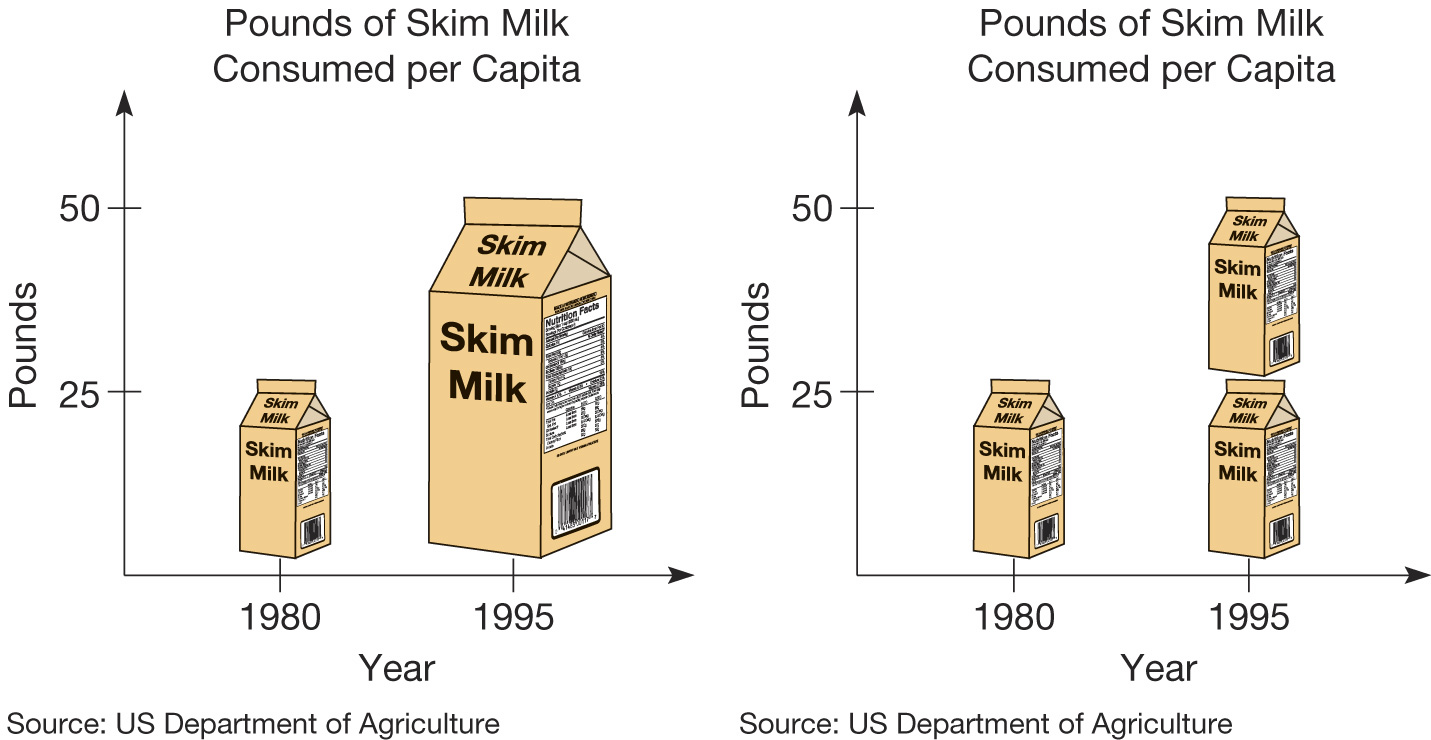


Here we have a totally irrelevant chart to answer the question.   
The answer would probably be a barplot citing in order the most excited phases of the Super Bowl game.

# **Wrong scaling**

  
  
To emphasize change over time it is acceptable to reduce the scale interval on the Y-axis.  
However, observing the first graph may falsely conclude to the fact that the unemployment rate was near 0 in 2000. This can be corrected by using a better-balanced scaling that gives a more fulfilled impression.

# **Pictographs**

  
  
Although pictographs are a quick way to convey information, it is really easy to lead into false conclusions just by mishandling the size. In the first pictograph, we may conclude that in 1995 US citizens were consuming way more (uncertainly how much more) milk than 1980. While the truth is revealed clearer in the second pictograph, which gives us the exact amount of difference by using same size pictographs.