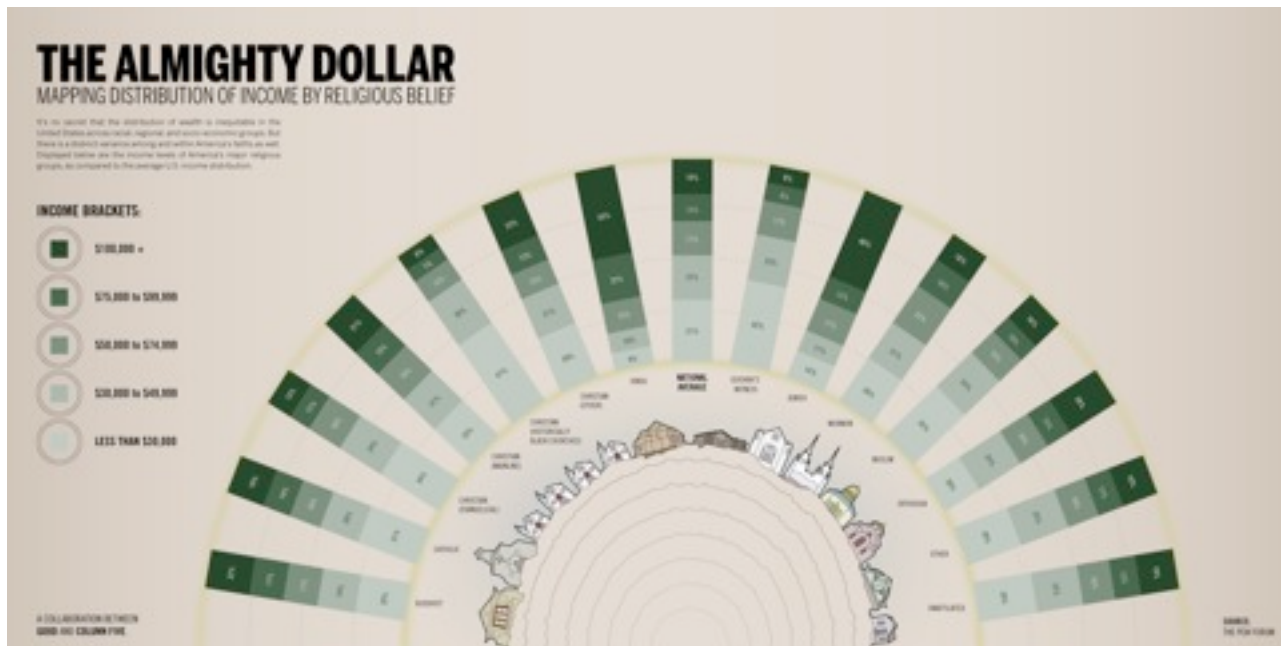


Stat645: Week 1

R revision and thinking about the grammar of graphics

The following graphic appeared in a recent post on good.is, an online magazine known for its data visualisations. Today we're going to dig into this plot, thinking about what it displays, how we can recreate, and explore variations that might display the data better.

Work in pairs to answer the following questions.



<http://awesome.good.is/transparency/web/1002/almighty-dollar/flat.html>

What did you learn from the plot that you didn't already know?

What data does the graphic show? Be as precise as you can.

Is there anything missing from the plot that might lead you to an incorrect conclusion?

Where did the data come from? Can you find the original source on the web?

I found the data, cleaned it up, and put it on the stat645 website. Download the data and open it up in R. How does it differ from the data on the plot?

Code sample 1 on the website creates a plot that is roughly equivalent to the good.is graphic. Describe what each of the three parts does.

Modify the code to remove the people who did not provide their income. What are the advantages and disadvantages of doing this?

Currently the plot focusses on the distribution of wealth given religion. Modify the code to look at the distribution of religion given wealth. What's the problem with displaying the data this way?

Stacked bar charts suffer from a perceptual problem - it's hard to read the heights of middle bars. Can you construct a display that uses points or lines to avoid this problem? (Hint: you may need to modify the data first)

What else could you do to improve this plot? Don't worry about how you would do it.