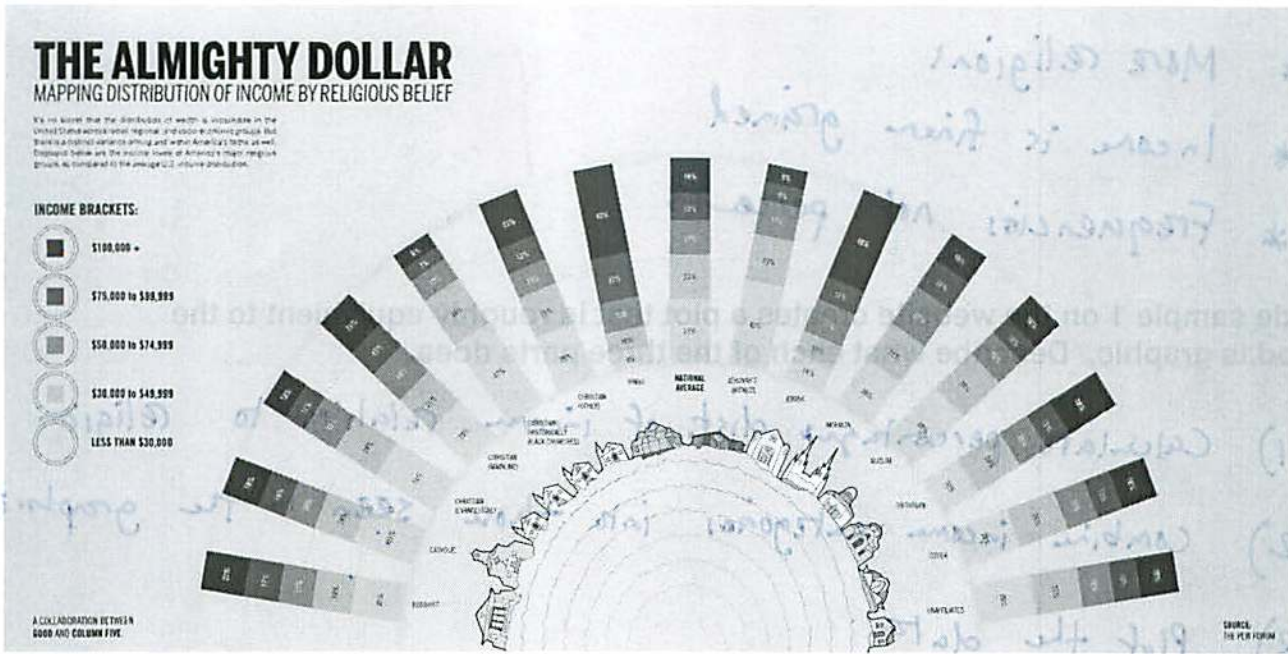


Stat645: Week 1

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?

Hindus & Jews are rich
Historically black churches & Jehovah's witnesses are poor

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

religion, income, and distribution (% of total) of incomes within religion

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

How many Hindus are there? Number of people within each religion is missing.

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

Pew Forum - <http://pewforum.org/Datasets/Dataset-Download.aspx>

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?

- * More religions
- * Income is finer grained
- * Frequencies not percents

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.

- 1) Calculate percentages dist. of income relative to religion
- 2) Combine income categories into those seen in the graphic
- 3) Plot the data
 - weight use prop.rel instead of counts
 - coord_flip() rotates 90° to make labels easier to read

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

```
relinc <- subset(relinc, !is.na(income2))  
# rerun step 1
```

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

- * Summarize proportions by income
- * Exchange religion & income on previous plot & switch weights

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)

See code online

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

- Reorder by average income

⋮