CSC 120 Group Project

1. Choose a dataset to analyze. Which dataset you choose to explore is up to you. You could find one on Kaggle or Gapminder, or, if group members are comfortable sharing data, you could use one or more members' requested personal data. For loads more options, see here. Not all datasets are equally suitable for analysis, so we will have brief (15 minute) group meetings to discuss in Week 10 (on 11/4 and, if necessary, 11/2). Come with at least a couple of ideas, and **do not share personal data before the meeting** (it's not always immediately clear whether sensitive information is contained in the data, so you should vet with me before sharing).
2. Produce one or more visualizations using `ggplot2`. The best visualizations tell a story about the data. Quality is much more important than quantity — one really interesting visualization is far preferable to ten visualizations that don't really say anything.
3. Fit a linear regression model. The best models will say something non-obvious about the data. For example, you might discover a relationship between two variables that you wouldn't have guessed existed, or you might show the opposite: that there is no relationship when you would have guessed that there is one.
4. You will present your findings to the class during the last week of classes. We will do a second meeting shortly before then to discuss (most likely in Week 14). In addition to the presentation, you should submit your analysis in written form before finals week begins.