

Lab #2: Practical R

6 October 2016

1 Data wrangling

The full set of height data gathered for Lab #1 will be sent out to you as a CSV file. Contact the TAs if you do not receive an email with the data. Read the data into R. Inspect and clean up the data as you deem necessary.

2 Plotting

Use the data to generate a series of *publication quality* plots. Carefully consider all aspects of your plots.

1. Graphically represent the distribution of heights in the dataset. Explain why you chose the representation(s) that you used.
2. Graphically represent the distribution of ages in the dataset. Explain why you chose the representation(s) that you used.
3. Make a scatterplot of age versus height, that conveys all the data. Color points by self-reported gender.
4. Make separate plots of age versus height for each self-reported gender.
5. Prepare a bar chart showing how often each letter of the alphabet occurs as the first letter of the last name, coloring bars by self-reported ethnicity. Make sure your plot includes all letters, not just those that are represented in the dataset.

3 Statistics

1. From this data, can you conclude that the average heights of men and women are different? Show your work. Justify your choice of analysis and interpret the results.
2. From this data, can you conclude that men and women tend to sit in different sections of the room? Recall that each group came from a different section. Hint: we've only looked at comparing two groups, so you may want to aggregate the groups we have in one or more ways (e.g., center versus sides, or left versus right).

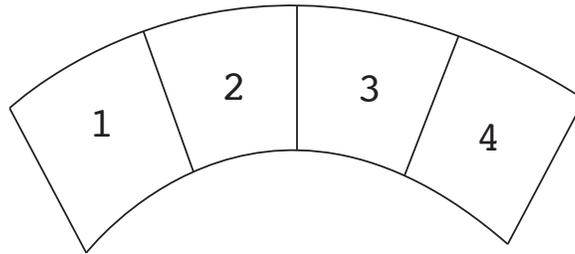


Figure 1: Section layout of C-200

Submit your answers, including plots and code by Thursday, 20 October 2016.