BIOS 611 Grading Report

Project 2

Student Name: Enying Gao

Summary

Item	Points
Completion	35
Code Presentation	10
Graphical Presentation	13
Interpretability	14
Statistical Correctness	9
Creativity	12
Total	93

Detailed Feedback

Completion

Students incorporated all the required tools and actions specified in the project definition.

> Points: 35 out of 35
>
> Comments: Because this project requires modular scripts, your .Rmd should only contain things for the report. Data analysis and plots should be a separate script, like your other R scripts.

Code Presentation

Formatting and quality of code organization and comments, e.g. code is clear, readable, and efficient with appropriate spacing, indentation, and comments.

> Points: 10 out of 10
>
> Comments: Clear and well-commented.

Graphical Presentation

Formatting and professionalism of figures/tables, e.g. axes on plots, variable names and units, sizes/shapes/colors, as well as titles and footnotes (if necessary) are defined.

> Points: 13 out of 15
>
> Comments: (-2) Plot 3: Year (x-axis) is not an integer. It does not make sense to mark 2012.5. It would be easier to read if you put HIC on one side and LMIC or color them differently. Plot 4: 'source_pm' in the title and text should be changed to words; no abbreviations or R dataset/variable name in report. In general, titles should describe what the plot is about, not

conclusions. Take plot 4 as an example, the title should be more like 'Percentages of different sources of particulate matter of human origin by year (1990-2015).

Interpretability

Quality of text description, e.g. well-written interpretation and conclusions are made instead of just plain plots and tables.

Points: 14 out of 15

Comments: (-1) Section 'Read Data': I can't tell what the four datasets are given you only have 3 .csv files and I can't find where the names (death, sdpm25, source_pm, air-quality) come from. You need to rename 'new_data', 'new_data1', new_data2'. Plot 3: Americas (HIC) do not have better air quality than LMIC either. It looks like it depends a lot on region.

Statistical Correctness

Is everything in the report (e.g. methods, outputs, visualization, conclusions) statistically correct?

Points: 9 out of 10

Comments: (-1) Section 'Figures': The print out of top 4 countries with the largest number of deaths seem to be countries with very high population. I think proportions of death over total pop, instead of number of deaths, would make more sense. Same for plot 1. Plot 2: I don't understand why gender would play a role in the number of deaths related to air pollution; could have looked into other things.

Creativity and Significance

Are there interesting and creative methods/outputs/conclusions? We encourage you to go above and beyond just reporting summary statistics.

Points: 12 out of 15

Comments: Interesting topic and data!