NYC Data Science Bootcamp

Exploratory Visualization Project

Introduction

One of the most important skills of any budding data scientist is the ability to present, communicate, and deliver their insights in a tangible manner that will be accessible to a wider audience. Often, good data visualization can transcend the barriers presented by both the technical and non-technical crowds -- but even the best visualization needs to tell a story. What story will you tell?

What We're Looking For

This project arguably has the quickest turnaround. You have just been thrown into the whirlwind of the bootcamp, haven't even learned your fellow students' names yet -- let alone the syntax of ggplot2 -- and you already have a project on the horizon? How are you possibly going to get it all done!?

The bottom line is: **You will find a way.**

For this project, we are purposefully putting you under time pressure in order to condition you to the potential hardships you may encounter throughout the bootcamp, and in the real world. Sometimes, things assigned today were due yesterday. Often, it may seem as though there is too much to be done and not enough time in the day.

The bottom line is: **You will still find a way.**

Needless to say, preparation will be key. Successful projects will encompass a plethora of skills including, but not limited to, the following:

- Submission in respect to the deadline.
- Background knowledge of dataset(s).

- Communication of motivation: why do we care?
- Research questions of interest: what do you want to find out?
- Answers to research questions: what have you uncovered?
- Presentation skills.
- Time management (not going over the allotted time).
- Ability to answer audience questions effectively and efficiently.
- Balance of complexity and simplicity.
- Explanation of future work: what would you do if given more time, data, etc.?
- Use of R and ggplot2.

The Details

Your project proposal declaration is due uniformly by the beginning of the Pulse Check on **Friday, April 15**. **No exceptions.** You must declare your dataset on the project proposal document and give a short background on some initial research questions. A sentence or two will suffice. these may change completely as you proceed with your analysis -- this is ok.

This is an **individual project** in respect to the final deliverable. **No exceptions.** Every student must have their own project and presentation; however, please feel free to collaborate and help each other with coding problems, insights, brainstorming, etc. We welcome this!

All code, data, etc. used to generate your graphics and any slides, markdown files, etc. intended for your presentation is due to the project GitHub repository uniformly by **Sunday, April 24 at 11:59pm**. **No exceptions.**

You will be required to deliver a **10 minute presentation** and respond to any audience questions. Time slots will be randomly assigned on this calendar, so all projects must be submitted on time. **No exceptions.**

An associated blog post will be due by **Sunday, May 1 at 11:59pm**. **No exceptions.** Remember, this is a living and breathing document. You may continue to develop and edit your project far beyond the deadline, as no project will ever truly be complete.

For inspiration, take a look at our previous students' blog posts or check out freely available datasets on NYC Open Data and the UCI Machine Learning Repository. There are so many resources out there -- a quick Google search for your topic of interest can be very fruitful!

For any lingering questions, please do not hesitate to reach out; we are always here to help!

Good luck!