PROJECT PLAN

Miriam Schoenbaum

October 25, 2023

1. List the datasets you plan to use for your project. Note the source of each dataset (dataMontgomery, Opportunity Atlas, census bureau, mentor organization, etc.) and give a link to it. Depending on your project topic and how large your datasets are, your list will probably be around 2-3 datasets.

* Data Montgomery
  + Non-motorist data: <https://data.montgomerycountymd.gov/Public-Safety/Crash-Reporting-Non-Motorists-Data/n7fk-dce5>
  + Driver data: <https://data.montgomerycountymd.gov/Public-Safety/Crash-Reporting-Drivers-Data/mmzv-x632>
  + Incident data: <https://data.montgomerycountymd.gov/Public-Safety/Crash-Reporting-Incidents-Data/bhju-22kf>
* Maryland Crash Data Download non-motorist data: <https://mdsp.maryland.gov/Pages/Dashboards/CrashDataDownload.aspx>
* The datasets will be combined into an analysis database consisting of the Non-motorist data, plus date of birth and crash date from the Maryland Crash Data Download, plus speed limit from the Driver data, plus hit & run and intersection type from the Incident data. Each row will represent one non-motorist hit.

1. List the questions you plan to research about your datasets. For each question, note how you plan to find the answer (visualization, t-test, k-means clustering, random simulation, bootstrapping, etc.) and which datasets you will use to find the answer. You should have 8 to 10 questions that you would like to answer across all of your datasets.

* Under what circumstances do drivers hit non-motorists aged 18 and under in Montgomery County?
  + location types – intersection vs. non-intersection, on-road vs. off-road, state vs. county vs. municipal, traffic control
  + timing – time of day, weekday vs. weekend (maybe), month
  + injury severity
  + crash report information – pedestrian movement, pedestrian actions, pedestrian location, pedestrian obeyed traffic signal, pedestrian at fault, hit & run
  + Answer plan: visualization (tables, charts)
* Are the circumstances different for non-motorists aged 18 and under vs. non-motorists aged older than 18?
  + location types – intersection vs. non-intersection, on-road vs. off-road, state vs. county vs. municipal, traffic control
  + timing – time of day, weekday vs. weekend (maybe), month
  + injury severity
  + crash report information – pedestrian movement, pedestrian actions, pedestrian location, pedestrian obeyed traffic signal, pedestrian at fault, hit & run
  + Answer plan: chi-squared tests, logistic regression
* In which locations do drivers hit non-motorists aged 18 and under in Montgomery County?
  + municipalities
  + County Council districts
  + proximity to schools
  + Answer plan: data visualization (mapping in ArcGIS)
* Are the locations different for non-motorists aged 18 and under vs. non-motorists aged older than 18?
  + municipalities
  + County Council districts
  + proximity to schools
  + Answer plan: chi-squared tests using data mapped in ArcGIS

1. Give the tools you plan to use for each phase of the project:
   1. Choice of Datasets & Topics: me!
   2. Data Ingestion & Wrangling (Cleaning): R
   3. EDA: R and ArcGIS
   4. Statistical Analysis: R, ArcGIS
   5. Data Visualization: R, ArcGIS, Excel
2. Think about the risks you face in completing your plan. Are there known unknowns you need to investigate? Contemplate unknown unknowns. It is hard to think of these by their very nature of being unknown, but use your imagination. Also, be honest and realist with yourself. Do you have other constraints on your time or resources needed to complete the project? How can you mitigate these risks.

* Known unknowns: The major one is crash location (latitude/longitude) in the police crash reports. It’s possible to correct the locations, but I don’t have the time or resources for that. Also the date/time format in the Data Montgomery dataset is terrible.
* Unknown unknowns: I’m very familiar with the Data Montgomery datasets but haven’t worked much with the Maryland Crash Data data.
* Other constraints on time and resources: YES!!!
* Risk mitigation: I may have to take a few days off work.