BI/Analytics Project

Deliverables

Week 7:

1. Project Proposal
   1. Together
2. Design Document
   1. Brenden
3. Ensure installation of R is possible
   1. McKay
4. Risks
   1. McKay

Week 8:

1. 2 cleaning methods each
   1. Filtering (NA’s etc)
      1. Brenden
   2. Remove unnecessary columns
      1. Brenden
   3. Editing new data for merge
      1. McKay
   4. Joins
      1. McKay
2. 1 profiling example each
   1. Summary table year and month
      1. Brenden
   2. Summary table intent and race when only homicide or suicide
      1. McKay
3. Installing R and RStudio
4. Learning basics of R
5. Pros and cons of cleaning and profiling in R
   1. McKay
6. List of things we learned

Week 9:

1. Write up pros and cons of R to Excel
2. 2 patterns/visualizations/outliers each
   1. Pairs plots
      1. McKay
   2. Boxplots
      1. McKay
   3. Bar-charts
      1. Brenden
   4. Outliers
      1. Brenden
3. Pros and cons
4. List of things we learned

Week 10:

1. Pros and cons of R
2. List of things we learned
3. 1 optimizing/forecasting/predicting model each
   1. Trend prediction (shootings increasing by year in certain areas/races)
      1. Brenden
   2. Full regression analysis
      1. McKay

Description of Problem

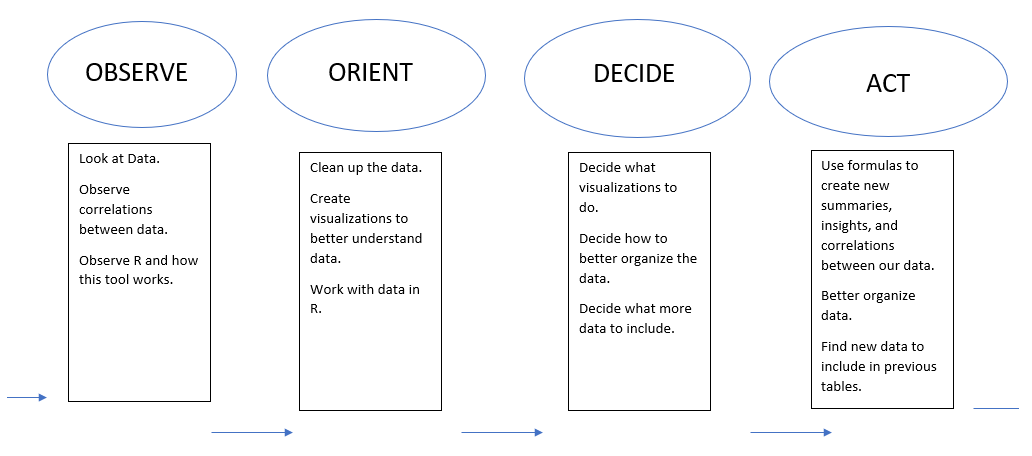
With the hot topic of gun related crimes, we wanted to look at variables that contribute to a person being killed by a gun. We were specifically curious to look at school shootings, race fueled shootings, age, gender, etc.

Dataset

<https://www.kaggle.com/hakabuk/gun-deaths-in-the-us>

Risks

1. New team that has never worked together before
   1. Team meetings at least once per week
2. New software to one team member
   1. McKay will help Brenden learn skills in R to complete this analysis while building his understanding of packages required for this analysis.
3. Limitations of data on types of models possible
   1. Searching for more supplementary data that may allow for other types of models

OODA Design

Week 8

**Brenden (R tutorials):**

**McKay (Regression research and Pros/Cons):**

1. Regression Research
   1. We are probably going to need to run a chi-squared test as the y variable is likely going to be categorical and non-binary. <https://stats.idre.ucla.edu/other/mult-pkg/whatstat/>
2. Pros and Cons of R to Excel for cleaning and profiling
   1. Pros
      1. R code is modular and syntactically simple
      2. R is capable of handling larger data sets with far greater speed
      3. The user cannot physically touch the data and only touches states of the data, so there is not much room for accidentally deleting important things in the data.
      4. R has many more tools available to it than Excel.
      5. R is open source.
      6. Streamlined data merges.
   2. Cons
      1. Data cannot be directly modified in its tabular form. This is difficult for beginners.
      2. Pivot tables don’t exist (although similar summaries of any kind can be created with relative ease).
      3. Excel is initially easier to use as most people have been exposed to similar Microsoft UI’s since they were five or from its inception.

**Week 8 Description of What We Did:**

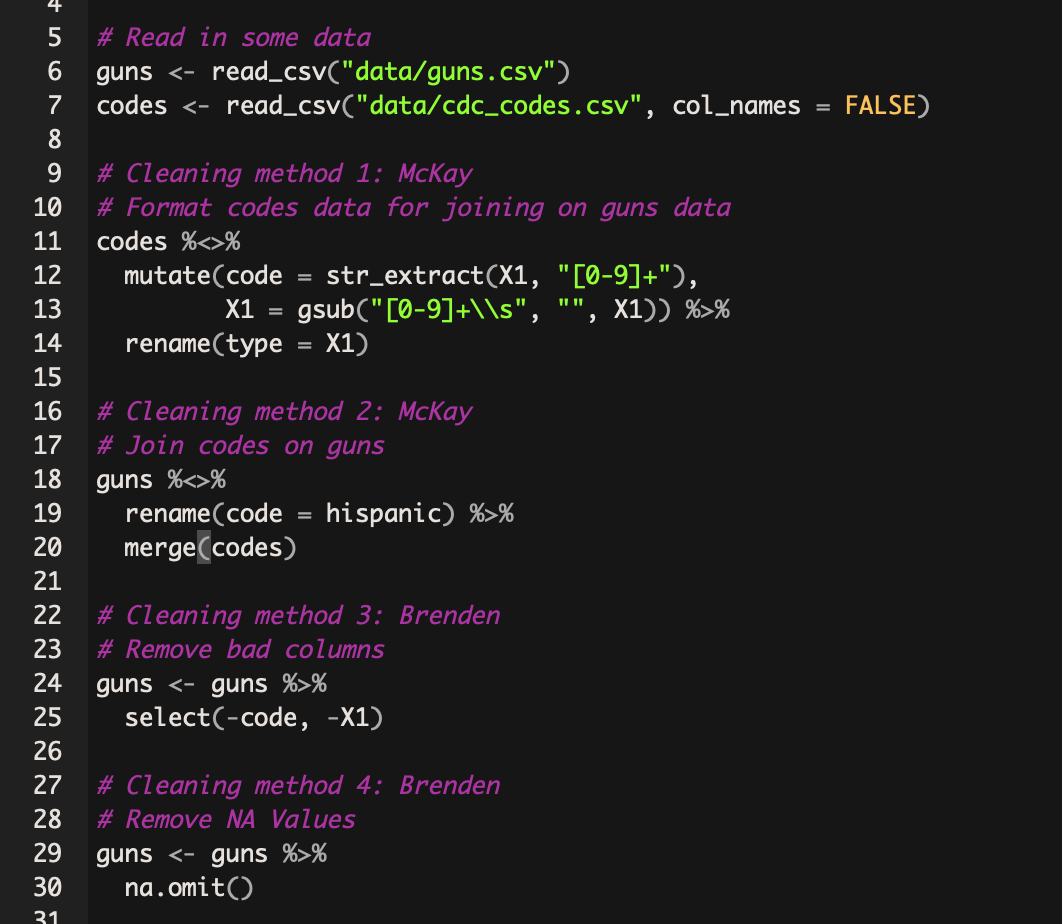
Edits to Project Proposal:

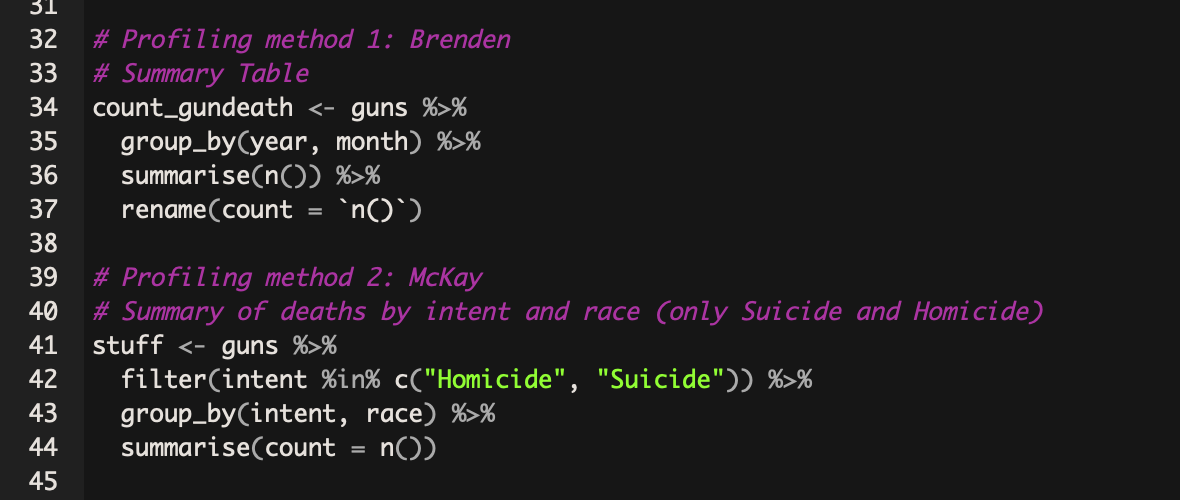
1. Changed cleaning methods and profiling to reflect what was needed
2. Edited pros/cons to be McKay’s task (we still did some together)

Methods:

We used the tidyverse package in R for reading in the data and cleaning it. We used different methods included in the package. Below is a selection of our code to illustrate our methods:

Reading in and cleaning the data:





Profiling:

The outputs: