Report 1

Hannah Lee, Khyati Jariwala, Michelle Xu

2022-10-21

Introduction

This city of Seattle probably makes you think about a few things: Starbucks, rainy weather, and popular tourist attractions, like the Space Needle. This prompted us to explore what more we can uncover about Seattle using seemingly unrelated datasets regarding weather, collisions, and population metrics to examine potential relationships between these variables.

The datasets we will be using are `Observed Monthly Rain Gauge Accumulations - Oct 2002 to May 2017`, which records monthly accumulations of rain gauges located throughout Seattle city limits, the `SDOT Collisions - All Years`, which records the number of collisions in Seattle (provided by SPD and recorded by Traffic Records), and the `City Annual Stats`, which includes the total population (and change in population), housing, and jobs for the City of Seattle for each calendar year. These datasets were retrieved from *data-seattlecitygis.opendata.arcgis.com* and *data.seattle.gov*, which are the city government's open access databases

These datasets are of interest to us, because we want to examine whether seemingly different variables, such as the weather, collisions, and population metrics can have potential relationships with each other. For the purpose of simplicity, we have selected these metrics and data for the city of Seattle rather than a whole country.

Unique rows in the `Observed Monthly Rain Gauge Accumulations - Oct 2002 to May 2017` dataset represent a monthly measurement of the rainwater accumulated in millimeters across various rain gauges located in Seattle. In the `SDOT Collisions - All Years` dataset, unique rows represent a detailed report of a collision occurring in Seattle with information about the number of people involved, the location of the collision, the severity of the collision, and other relevant information. For the `City Annual Stats` dataset, the unique rows are representative of annual population, housing, job, and industry data for the city of Seattle.

These datasets will be reshaped and tidied individually first, and then joined by the variable *Year*, which is the common key between all of our datasets. Since two datasets contain month and day data, those will be joined by the *Month*, *Day*, and *Year* keys prior to being joined with the City Annual Stats dataset, which only has the *Year* key. All datasets are predominantly numeric (rainfall, population, jobs, etc.); however, there are a few categorical variables interspersed (year, collision severity, etc.).

A potential trend we expect to see once we combine the datasets is that the number of collisions should increase as the average rainfall increases. This is because rain causes roads to be slippery, which could result in collisions. Additionally, the more rainfall there is, the severity of the collision should also increase for the same reason. For population and collisions, we expect that with an increasing population the number of collisions would increase, because there would be more people on the road. Within the City Annual Stats dataset, we expect to see that the number of jobs will increase as the total population increases due to the fact that people are attracted to live in cities with more job opportunities, because people need to be able to make a living.

Tidying

```r
# Load libraries
library(tidyverse)

# Read in datasets
collisions <- read.csv("Collisions.csv") %>% as.data.frame()
citystats <- read.csv("City_Annual_Stats.csv") %>% as.data.frame()
rainfall <- read.csv("Observed_Monthly_Rain_Gauge_Accumulations_-_Oct_2002_to_May_2017.csv") %>% as.data

# viewing datasets
collisions %>% head()
```

```
##           X        Y OBJECTID INCKEY COLDETKEY REPORTNO    STATUS ADDRTYPE
## 1 -122.3805 47.67528        1 352527    353987  EC56520 Unmatched    Block
## 2 -122.3142 47.66750        2   1305      1305  3502004   Matched    Block
##   INTKEY                                    LOCATION EXCEPTRSNCODE
## 1     NA    NW 64TH ST BETWEEN 17TH AVE NW AND 20TH AVE NW
## 2     NA BROOKLYN AVE NE BETWEEN NE 52ND ST AND NE 55TH ST
##   EXCEPTRSNDESC SEVERITYCODE                   SEVERITYDESC COLLISIONTYPE
## 1                          1 Property Damage Only Collision
## 2                          2                Injury Collision     Left Turn
##   PERSONCOUNT PEDCOUNT PEDCYLCOUNT VEHCOUNT INJURIES SERIOUSINJURIES FATALITIES
## 1           2        0           0        0        0               0          0
## 2           2        0           0        2        1               0          0
##                  INCDATE              INCDTTM
## 1 2022/06/16 00:00:00+00            6/16/2022
## 2 2013/03/27 00:00:00+00 3/27/2013 9:15:00 AM
##                               JUNCTIONTYPE SDOT_COLCODE
## 1 Mid-Block (not related to intersection)           14
## 2                        Driveway Junction          11
##                                            SDOT_COLDESC INATTENTIONIND
## 1           MOTOR VEHICLE STRUCK MOTOR VEHICLE, REAR END
## 2 MOTOR VEHICLE STRUCK MOTOR VEHICLE, FRONT END AT ANGLE
##   UNDERINFL WEATHER ROADCOND LIGHTCOND PEDROWNOTGRNT SDOTCOLNUM SPEEDING
## 1                                                            NA
## 2         N   Clear      Dry  Daylight                       NA
##   ST_COLCODE                                              ST_COLDESC SEGLANEKEY
## 1         NA                                                                  0
## 2         28 From opposite direction - one left turn - one straight          0
##   CROSSWALKKEY HITPARKEDCAR
## 1            0            Y
## 2            0            N
##  [ reached 'max' / getOption("max.print") -- omitted 4 rows ]
```

```r
citystats %>% head()
```

```
##      City Year Const_Res  FIRE Manufacturing Retail Services   WTU Government
## 1 Seattle 1995     15282 35253         38050  31504   185899 40545      51571
## 2 Seattle 2000     22645 42471         37104  41984   235336 43636      47565
## 3 Seattle 2001     21601 41671         35044  42232   234726 42056      48104
## 4 Seattle 2002     19582 40710         31094  38534   219499 37943      47518
## 5 Seattle 2003     17831 41005         28425  37179   217129 39494      48424
##   Education Total_Jobs Housing_Units Total_Population Households Year_Display
## 1     28625     426729        259864           541509     250050         1995
```

```
## 2     32094     502835        270524           563376     258499         2000
## 3     31771     497205        273651           565228     259691         2001
## 4     33882     468763        277905           568908     261767         2002
## 5     32723     462210        280969           572472     263791         2003
##   Change_Population Change_Housing_Units Change_Jobs ObjectID
## 1                NA                   NA          NA        1
## 2                NA                   NA          NA        2
## 3              1852                 3127       -5630        3
## 4              3680                 4254      -28442        4
## 5              3564                 3064       -6553        5
##  [ reached 'max' / getOption("max.print") -- omitted 1 rows ]
```

```
rainfall %>% head()
```

```
##         Date RG01 RG02 RG03 RG04 RG05 RG07 RG08 RG09 RG10_30 RG11 RG12 RG14
## 1 11/30/2002 2.43 3.36 2.88 2.48 0.78 2.49 2.57 2.93    3.25 2.38 2.59 2.46
## 2 12/31/2002 4.31 1.40 5.46 4.80 1.99 5.06 2.48 2.35    6.48 4.95 5.71 3.57
## 3 01/31/2003 6.55 7.35 5.84 6.48 7.57 4.47 7.39 7.31    5.42 6.58 7.58 5.72
## 4 02/28/2003 1.61 1.81 1.70 1.49 1.11 1.50 1.56 1.73    1.18 1.37 1.47 1.33
## 5 03/31/2003 5.01 5.88 3.12 5.01 5.09 5.15 5.14 5.01    5.68 4.01 5.16 4.57
##   RG15 RG16 RG17 RG18 RG20_25
## 1 3.06 2.69 3.59 3.17    3.15
## 2 5.77 3.28 5.77 6.02    5.60
## 3 7.47 8.32 9.69 7.66    7.17
## 4 1.19 1.21 1.52 1.09    1.34
## 5 5.50 5.61 5.62 5.49    4.89
##  [ reached 'max' / getOption("max.print") -- omitted 1 rows ]
```

```
# Tidying "Collisions" dataset
collisions %>%
  group_by(INCDTTM, COLLISIONTYPE, SEVERITYDESC, WEATHER, VEHCOUNT) %>% # there are columns we wanted t
  summarise(n = n()) %>% # we reduced this exceedingly large dataset by transforming it around a differ
  separate(INCDTTM, into = c("Month", "Day", "Year")) %>%
  mutate_at(c("Year", "Month", "Day"), as.integer) %>%
  filter(COLLISIONTYPE != "") %>%
  arrange(n) %>%
  ungroup() -> collisions_cleaned
```

```
# create "working" dataset by keep only useful columns of dataset
citystats %>%
  select(-c(City, Year_Display, ObjectID, Change_Population, Change_Housing_Units, Change_Jobs)) -> city
```

```
# Tidying rainfall dataset
rainfall %>%
  pivot_longer(cols = c("RG01":"RG20_25"),
               names_to = "Gauge_Location",
               names_transform = as.factor,
               values_to = "Accumulated_Rainfall",
               values_transform = as.numeric) %>%
  separate(Date, into = c("Month", "Day", "Year")) %>%
  mutate_at(c("Year", "Month", "Day"), as.integer) %>%
  group_by(Year, Month) %>%
  summarise(Rain_Accum = mean(Accumulated_Rainfall)) %>%
  ungroup()-> rain_cleaned # we think its in in
```

To begin, we modified two of our datasets so that they are tidy. For the `Observed Monthly Rain Gauge`

`Accumulations - Oct 2002 to May 2017` dataset, we created a new variable for *Gauge_Location* and for *Accumulated_Rainfall*, which contains the corresponding amount of rainfall accumulated for each location. Both the `Observed Monthly Rain Gauge Accumulations - Oct 2002 to May 2017` and `SDOT Collisions - All Years` datasets provided dates in the format of MM/DD/YY. Our goal is to join all of our datasets by year, so we separated the given dates into variables for month, day, and year.

Joining/Merging

The total number of observations and unique IDs in the `citystats`, `rain_cleaned` and `collisions_cleaned` datasets are 25;19, 175;3, 193545;8, respectively. The only ID that all of the datasets have in common is *Year*. While both `rain_cleaned` and `collisions_cleaned` have a *Month* ID, only `collisions_cleaned` has a *Day* ID. The unique IDs that appear in only the `citystats` dataset include *City*, *Const_Res*, *FIRE*, *Manufacturing*, *Retail*, *Services*, *WTU*, *Government*, *Education*, *Total_Jobs*, *Housing_Units*, *Total_Population*, *Households*, *Year_Display*, *Change_Population*, *Change_Jobs*, and *ObjectID*. The unique ID that appears only in the `rain_cleaned` dataset is *Rain_Accum*. Lastly, the unique IDs that appear only in the `collisions_cleaned` dataset are *COLLISIONTYPE*, *SEVERITYDESC*, *WEATHER*, *VEHCOUNT*, and *n*. There were no IDs that have been left out or any rows that were dropped/added while joining the datasets. Note that we removed a few columns that did not seem to provide meaningful information (particularly from the `citystats` dataset).

```r
# find total number of observations for each dataset and unique IDs
dim(city_cleaned)
```

```
## [1] 25 13
```

```r
dim(rain_cleaned)
```

```
## [1] 175   3
```

```r
dim(collisions_cleaned)
```

```
## [1] 193545      8
```

```r
# joining cleaned data sets, where each row now represents a collision incident
collisions_cleaned %>%
  left_join(rain_cleaned, by = c("Year", "Month")) %>%
  left_join(city_cleaned, by = "Year") -> city_rain_collisions

# finding dimensions of & viewing joint dataset
dim(city_rain_collisions)
```

```
## [1] 193545     21
```

```r
city_rain_collisions
```

```
## # A tibble: 193,545 x 21
##    Month   Day  Year COLLI~1 SEVER~2 WEATHER VEHCO~3     n Rain_~4 Const~5  FIRE
##    <int> <int> <int> <chr>   <chr>   <chr>     <int> <int>   <dbl>   <int> <int>
##  1     1     1  2004 Angles  Injury~ Snowing       2     1     5.5   18157 40063
##  2     1     1  2004 Angles  Proper~ Clear         2     1     5.5   18157 40063
##  3     1     1  2004 Angles  Seriou~ Raining       2     1     5.5   18157 40063
##  4     1     1  2004 Left T~ Injury~ Overca~       2     1     5.5   18157 40063
##  5     1     1  2004 Left T~ Injury~ Raining       2     1     5.5   18157 40063
##  6     1     1  2004 Other   Injury~ Clear         2     1     5.5   18157 40063
##  7     1     1  2004 Other   Proper~ Raining       2     1     5.5   18157 40063
##  8     1     1  2004 Other   Proper~ Snowing       1     1     5.5   18157 40063
##  9     1     1  2004 Parked~ Proper~ Clear         3     1     5.5   18157 40063
## 10     1     1  2004 Parked~ Proper~ Overca~       2     1     5.5   18157 40063
## # ... with 193,535 more rows, 10 more variables: Manufacturing <int>,
```

```
## #   Retail <int>, Services <int>, WTU <int>, Government <int>, Education <int>,
## #   Total_Jobs <int>, Housing_Units <int>, Total_Population <int>,
## #   Households <int>, and abbreviated variable names 1: COLLISIONTYPE,
## #   2: SEVERITYDESC, 3: VEHCOUNT, 4: Rain_Accum, 5: Const_Res
```

Wrangling

After manipulating our dataset, we were able to compute and analyze summary statistics for many of our variables.

```
# viewing "working" dataset
city_rain_collisions
```

```
## # A tibble: 193,545 x 21
##    Month   Day  Year COLLI~1 SEVER~2 WEATHER VEHCO~3     n Rain_~4 Const~5  FIRE
##    <int> <int> <int> <chr>   <chr>   <chr>     <int> <int>   <dbl>   <int> <int>
##  1     1     1  2004 Angles  Injury~ Snowing       2     1     5.5   18157 40063
##  2     1     1  2004 Angles  Proper~ Clear         2     1     5.5   18157 40063
##  3     1     1  2004 Angles  Seriou~ Raining       2     1     5.5   18157 40063
##  4     1     1  2004 Left T~ Injury~ Overca~       2     1     5.5   18157 40063
##  5     1     1  2004 Left T~ Injury~ Raining       2     1     5.5   18157 40063
##  6     1     1  2004 Other   Injury~ Clear         2     1     5.5   18157 40063
##  7     1     1  2004 Other   Proper~ Raining       2     1     5.5   18157 40063
##  8     1     1  2004 Other   Proper~ Snowing       1     1     5.5   18157 40063
##  9     1     1  2004 Parked~ Proper~ Clear         3     1     5.5   18157 40063
## 10     1     1  2004 Parked~ Proper~ Overca~       2     1     5.5   18157 40063
## # ... with 193,535 more rows, 10 more variables: Manufacturing <int>,
## #   Retail <int>, Services <int>, WTU <int>, Government <int>, Education <int>,
## #   Total_Jobs <int>, Housing_Units <int>, Total_Population <int>,
## #   Households <int>, and abbreviated variable names 1: COLLISIONTYPE,
## #   2: SEVERITYDESC, 3: VEHCOUNT, 4: Rain_Accum, 5: Const_Res
```

```
# create table of summary statistics for numerical variables, such as rainfall, vehicles in collisions,
city_rain_collisions %>%
  select(Year, Rain_Accum, Total_Population, Total_Jobs, SEVERITYDESC, WEATHER, VEHCOUNT) %>%
  filter(Year %in% c(2002:2017)) %>%
  mutate(Ratio_jobs_per_pop = Total_Jobs / Total_Population) %>%
  ungroup() %>%
  na.omit() %>%
  summarise_if(is.numeric, list(mean = mean, sd = sd, min = min, max = max))
```

```
## # A tibble: 1 x 24
##   Year_mean Rain_Accum~1 Total~2 Total~3 VEHCO~4 Ratio~5 Year_sd Rain_~6 Total~7
##       <dbl>        <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     2010.         3.16 612586. 492522.    1.97   0.804    3.91    2.58  35604.
## # ... with 15 more variables: Total_Jobs_sd <dbl>, VEHCOUNT_sd <dbl>,
## #   Ratio_jobs_per_pop_sd <dbl>, Year_min <int>, Rain_Accum_min <dbl>,
## #   Total_Population_min <int>, Total_Jobs_min <int>, VEHCOUNT_min <int>,
## #   Ratio_jobs_per_pop_min <dbl>, Year_max <int>, Rain_Accum_max <dbl>,
## #   Total_Population_max <int>, Total_Jobs_max <int>, VEHCOUNT_max <int>,
## #   Ratio_jobs_per_pop_max <dbl>, and abbreviated variable names
## #   1: Rain_Accum_mean, 2: Total_Population_mean, 3: Total_Jobs_mean, ...
```

For the variable *VEHCOUNT*, the vehicles affected in a collision, has a mean of 1.961 vehicles and a standard deviation of 0.589 vehicles. Using the mutate dplyr function, we were able to find the ratio of total number of jobs to the total population is 0.803 on average with a standard deviation of 0.02. For the variable

Rain_Accum, we found that the monthly rain gauge accumulation is 5.3 inches on average with a standard deviation of 2.89 inches.

```
# create frequency tables for collision type and weather (summary statistics for categorical variables)
city_rain_collisions %>%
  select(COLLISIONTYPE) %>%
  group_by(COLLISIONTYPE) %>%
  summarise(Frequency = n()) %>%
  arrange(desc(Frequency))
```

```
## # A tibble: 10 x 2
##    COLLISIONTYPE Frequency
##    <chr>             <int>
##  1 Parked Car        45860
##  2 Angles            35659
##  3 Rear Ended        34238
##  4 Other             24885
##  5 Sideswipe         18799
##  6 Left Turn         14372
##  7 Pedestrian         8028
##  8 Cycles             6217
##  9 Right Turn         3117
## 10 Head On            2370
```

```
city_rain_collisions %>%
  select(WEATHER) %>%
  group_by(WEATHER) %>%
  summarise(Frequency = n()) %>%
  arrange(desc(Frequency))
```

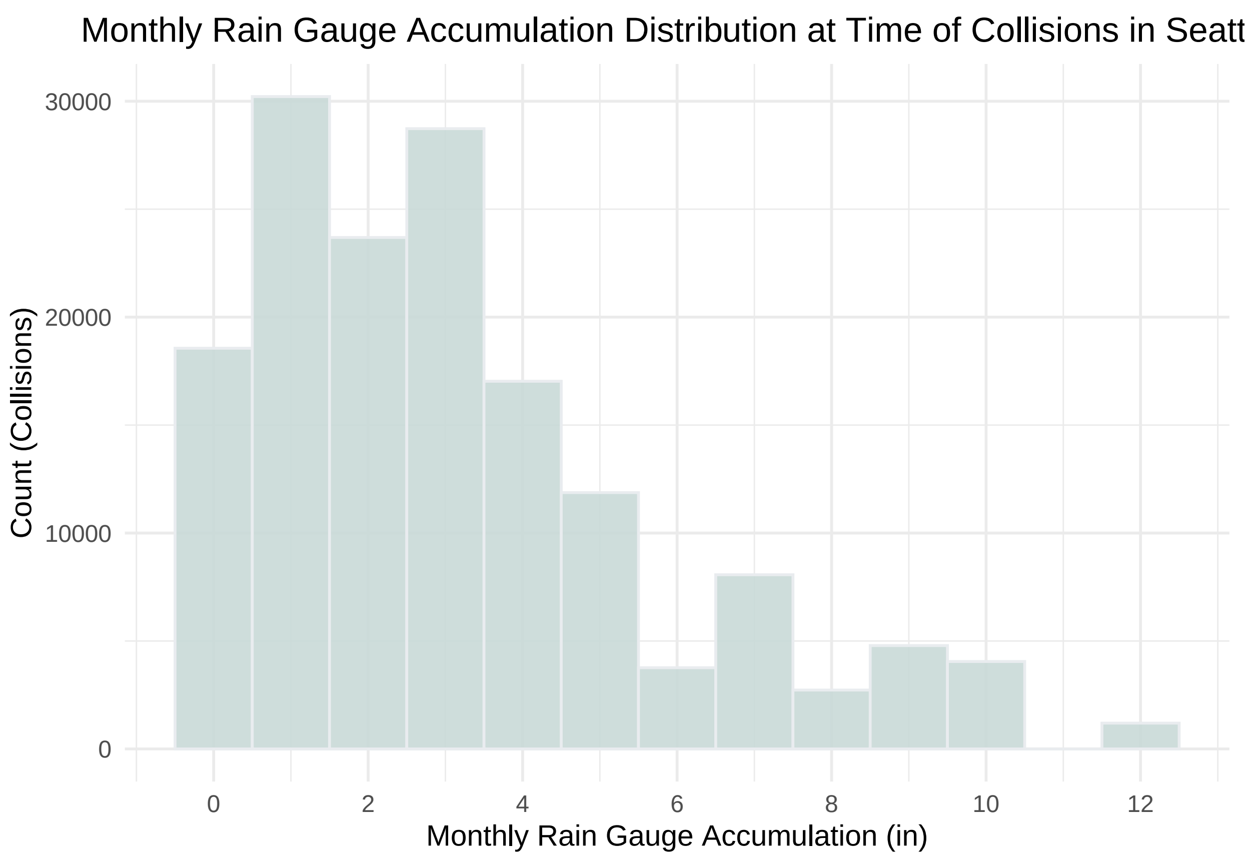
```
## # A tibble: 13 x 2
##    WEATHER                    Frequency
##    <chr>                          <int>
##  1 "Clear"                       113283
##  2 "Raining"                      33774
##  3 "Overcast"                     29427
##  4 "Unknown"                      13945
##  5 "Other"                         1010
##  6 "Snowing"                        943
##  7 "Fog/Smog/Smoke"                 671
##  8 ""                               247
##  9 "Sleet/Hail/Freezing Rain"       125
## 10 "Blowing Sand/Dirt"               57
## 11 "Partly Cloudy"                   32
## 12 "Severe Crosswind"                29
## 13 "Blowing Snow"                     2
```

Furthermore, we created a frequency table for the two categorical variables we were interested in, *COLLISIONTYPE* and *WEATHER*. These tables represent the number of collisions observed for each collision type and weather condition. In our dataset, we can see that the collisions most frequently occurred with parked cars and least frequently occurred when it was head-on. Additionally, in our dataset, the most collisions occurred when the weather was clear, and the least collisions occurred when there was blowing snow.

Visualizing

```
# prepare data for visualization
city_rain_collisions %>%
  mutate(Ratio_jobs_per_pop = Total_Jobs / Total_Population) -> df
```

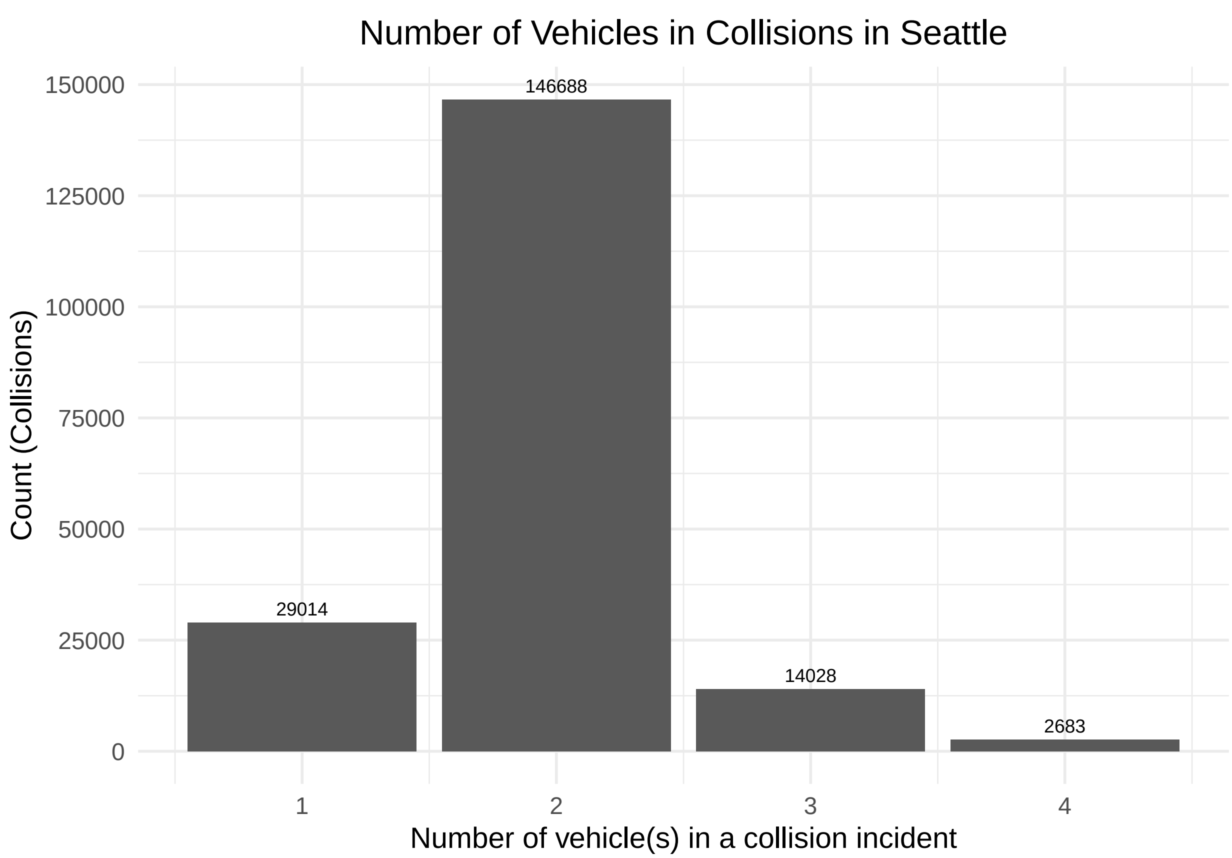
```
# rainfall histogram (1 variable)
df %>%
  ggplot(aes(x=Rain_Accum)) +
  geom_histogram(
    binwidth = 1,
    fill="#c9dad8",
    color="#e9ecef",
    alpha=0.9) +
  scale_x_continuous(breaks = seq(0, 12, 2)) +
  labs(x = "Monthly Rain Gauge Accumulation (in)", y = "Count (Collisions)", title = "Monthly Rain Gaug
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



The *Monthly Rain Gauge Accumulation Distribution at Time of Collisions in Seattle* histogram represents the distribution of the (average) rain gauge accumulation (inches) throughout the city of Seattle for the month of each observation, or row, in our dataset. This histogram is unimodal and right-skewed. Since each row in our dataset represents a collision, this histogram presents an unexpected trend that as the monthly rain gauge accumulation increases, the number of observed/reported collision counts decreases.

```
# vehicle count in collision incidents histogram (1 variable)
df %>%
```

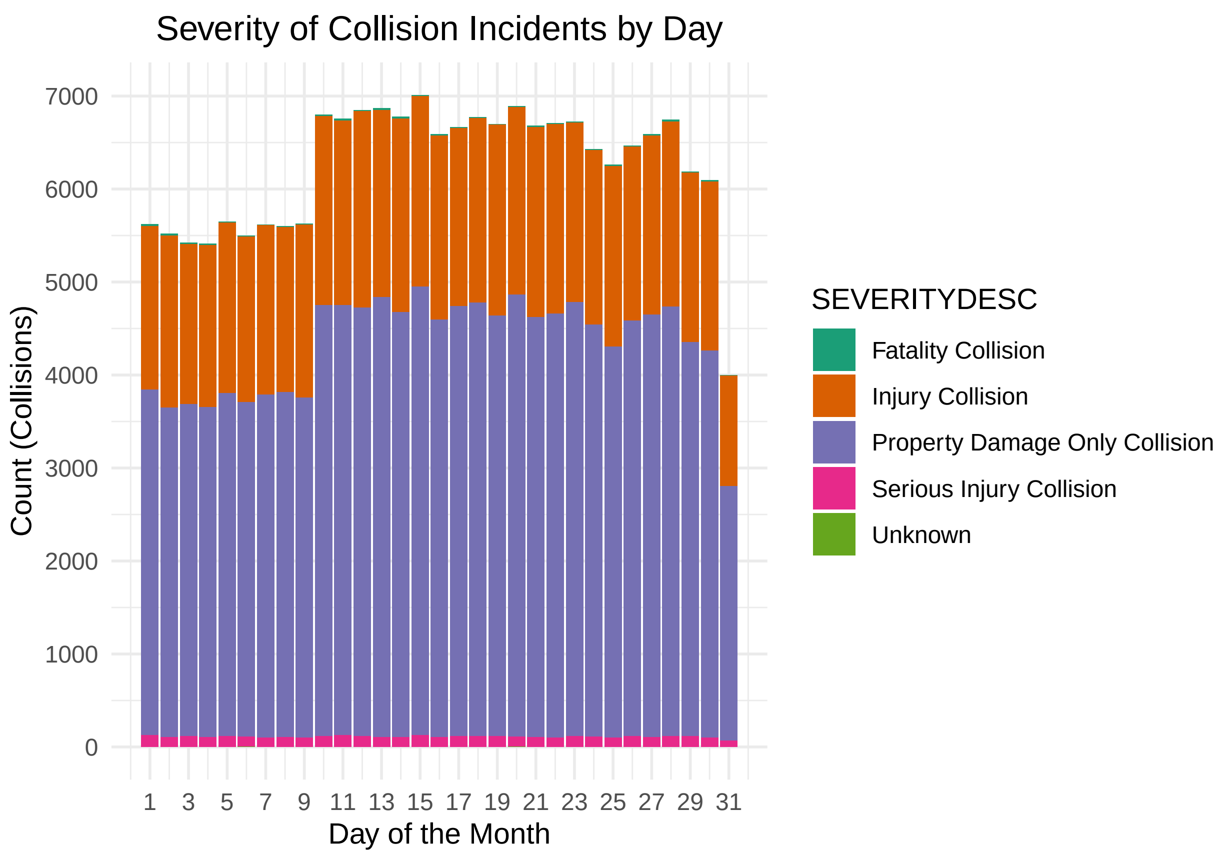
```r
  filter(VEHCOUNT %in% seq(1:4)) %>%
  ggplot(aes(x = VEHCOUNT)) +
  stat_count(geom = "bar") +
  geom_text(stat='count',
            aes(label=..count..),
            vjust = -0.5,
            size = 2.5) +
  scale_x_continuous(breaks = seq(1, 4, 1)) +
  scale_y_continuous(breaks = seq(0, 150000, 25000)) +
  labs(x = "Number of vehicle(s) in a collision incident",
       y = "Count (Collisions)",
       title = "Number of Vehicles in Collisions in Seattle") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



The *Number of Vehicles Involved in Seattle Collisions* bar graph represents the vehicle counts of the collisions of Seattle contained in our dataset. Note, we disregarded any vehicle counts in our dataset that were less than 1 or greater than 4. Clearly, the bar graph depicts that most of the collisions that had occurred affected two vehicles.

```r
# Collisions vs. Day by Severity (2 variables)
df %>%
  ggplot(aes(x=Day)) +
  geom_bar(aes(fill = SEVERITYDESC)) +
  scale_x_continuous(breaks = seq(1, 31, 2)) +
  scale_y_continuous(breaks = seq(0, 7000, 1000)) +
```

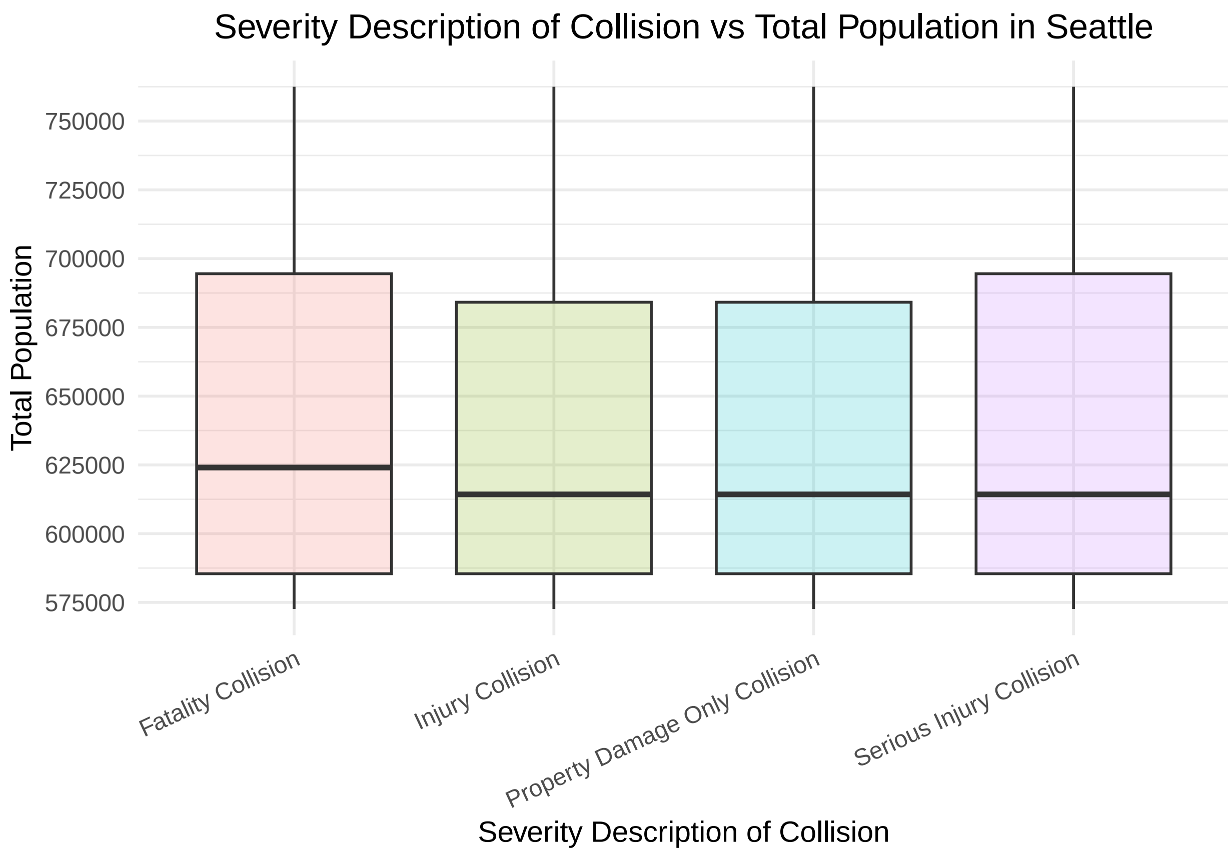
```
  scale_fill_brewer(palette = "Dark2") +
  labs(
    x = "Day of the Month",
    y = "Count (Collisions)",
    title = "Severity of Collision Incidents by Day"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



The *Severity of Collision Incidents by Day* bar plot represents the count (number of collisions) of each day of the month by severity of the collision. From the plot, we see that the property damage only collisions consistently occur the most over the days in the month. Similarly, injury collisions are the second most frequently occurring collisions.

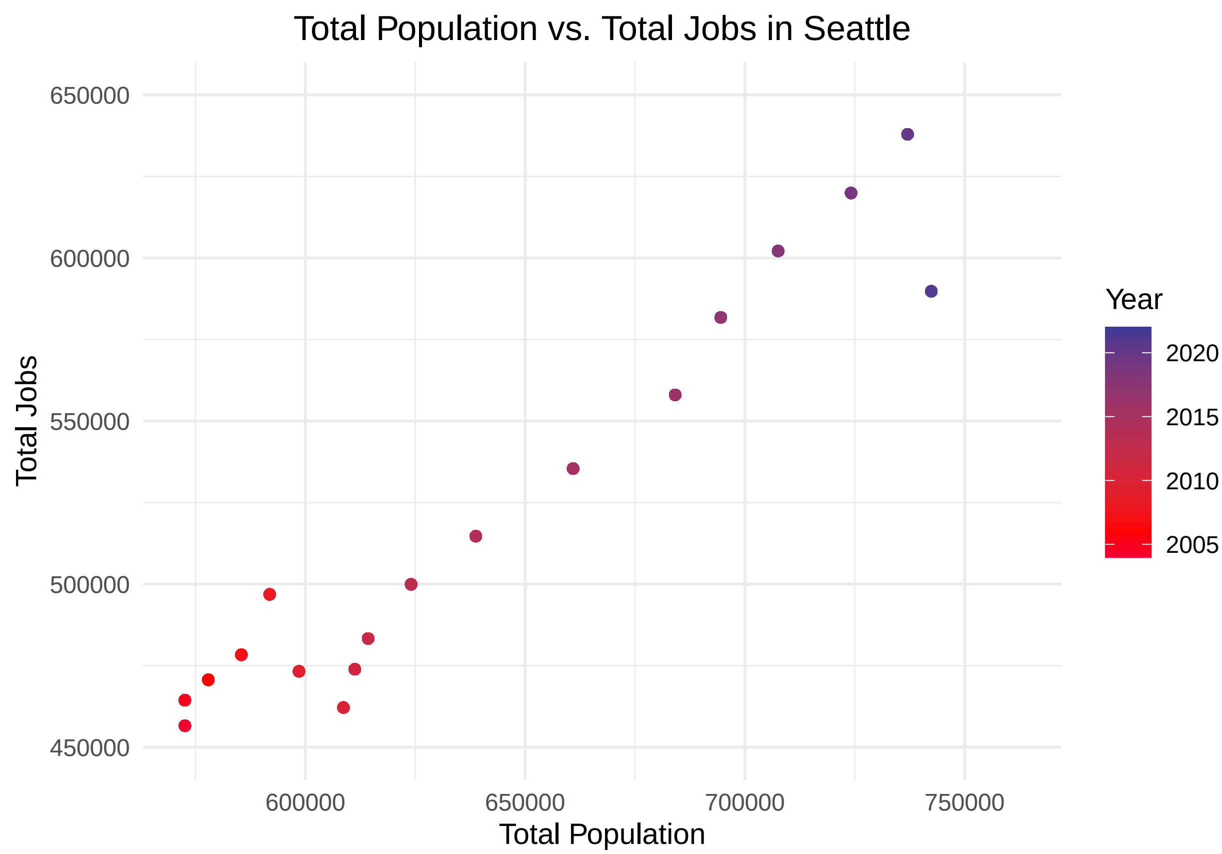
```
# Severity vs Population (2 variables)
df %>%
  filter(SEVERITYDESC != "Unknown") %>%
  ggplot(aes(x=SEVERITYDESC, y = Total_Population, fill = SEVERITYDESC)) +
  geom_boxplot(alpha = 0.2) +
  scale_y_continuous(breaks = seq(500000,800000,25000)) +
  labs(x = "Severity Description of Collision",
       y = "Total Population",
       title = "Severity Description of Collision vs Total Population in Seattle") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 25,
                                    hjust = 1),
```

```
        legend.position = "none",
        plot.title = element_text(hjust = 0.5))
```



The *Severity Description of Collision vs. Total Population in Seattle* boxplot represents the total population in Seattle at the time of each collision categorized by severity. Note that we excluded collisions that had an unknown severity description, as it did not seem to contribute in showing how the total population differs among varying collision severity. The plot indicates that the total spread of population does not clearly differ among different levels of severity descriptions of the collision. Thus, the boxplot does not support that a larger population indicates a higher number of severe collision occurrences compared to milder cases of collisions.

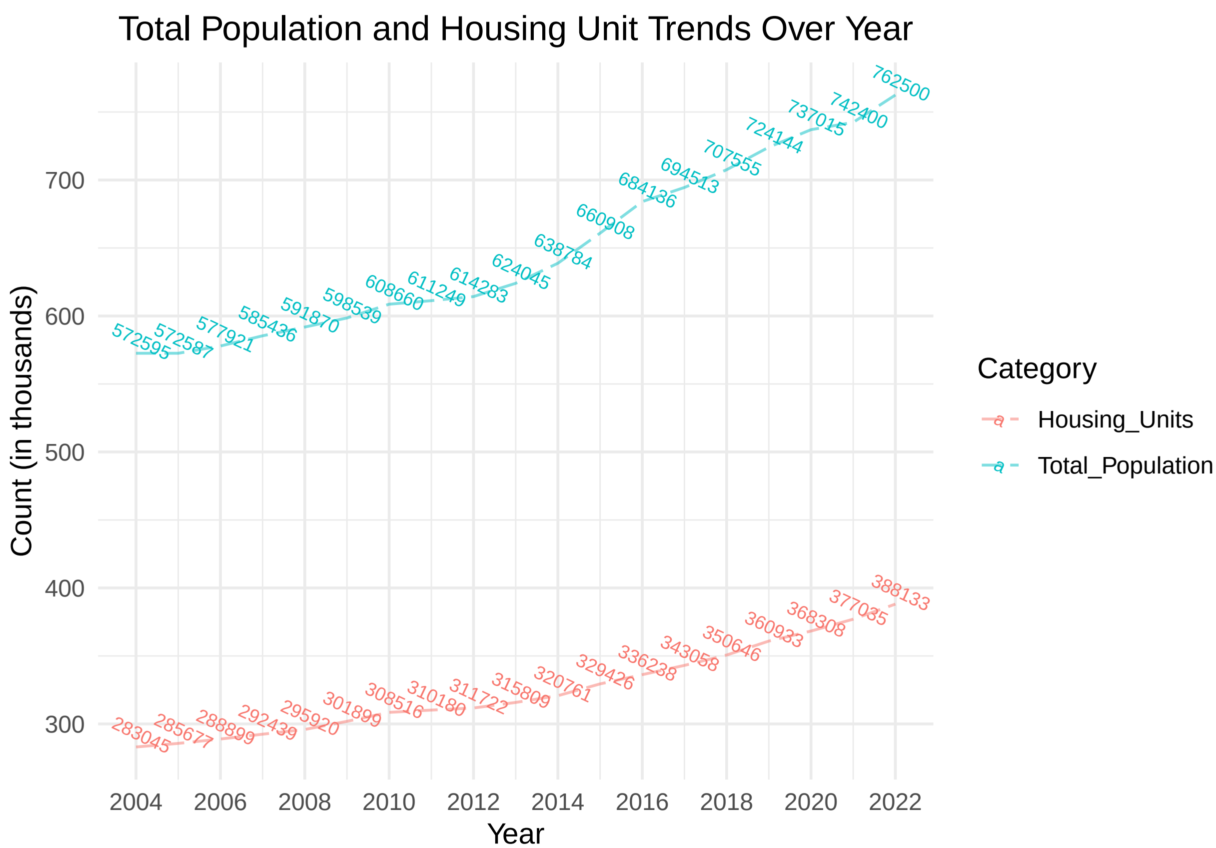
```
# Jobs vs. Population by Year (3 variables)
df %>%
  ggplot(aes(x=Total_Population, y = Total_Jobs)) +
  geom_point(aes(color = Year)) +
  scale_color_gradient2(low = "blue", mid = "red", midpoint = 2006) +
  labs(x = "Total Population", y = "Total Jobs", title = "Total Population vs. Total Jobs in Seattle") +
  ylim(450000,650000) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



The *Total Population vs. Total Jobs in Seattle* scatterplot represents the relationship between the total population and total number of jobs in Seattle by year. The plot suggests that there may be a positive linear relationship between total population and total number of jobs. That is, as the total population increases, the total number of jobs do as well. Furthermore, both the total population and number of jobs increases over the years.

```
# Population/Housing vs. Year (3 variables)
library(scales)
df %>%
  select(Year, Total_Population, Housing_Units) %>%
  distinct() %>%
  pivot_longer(
    cols = c("Total_Population", "Housing_Units"),
    names_to = "Category",
    values_to = "Stat"
  )  %>%
  ggplot(aes(x = Year, y = Stat)) +
  geom_line(aes(color = Category),
            stat = "identity",
            alpha = 0.5,
            linetype = "longdash") +
  geom_text(aes(label=Stat,
                color = Category),
            stat='identity',
            angle = -25,
            vjust = -0.5,
            size = 2.5) +
```

```
scale_x_continuous(breaks = seq(2004, 2022, 2)) +
scale_y_continuous(labels = label_number(scale = 1e-3)) +
labs(
  x = "Year",
  y = "Count (in thousands)",
  title = "Total Population and Housing Unit Trends Over Year") +
theme_minimal() +
theme( plot.title = element_text(hjust = 0.5) )
```



The *Total Population and Housing Unit Trends Over Year* plot depicts the total population and housing unit counts in Seattle over the years. Based on the plot, there is a trend of increasing total population and housing units as the years go by. Beyond this, total population and housing units seem to increase at very similar rates.

References

1) Observed Monthly Rain Gauge Accumulations - Oct 2002 to May 2017: This data source includes monthly accumulations of rain gauges located throughout Seattle city limits. https://data.seattle.gov/City-Business/Observed-Monthly-Rain-Gauge-Accumulations-Oct-2002/rdtp-hzy3
2) SDOT Collisions - All Years: This data source includes records of collisions in Seattle (provided by SPD and recorded by Traffic Records). https://data-seattlecitygis.opendata.arcgis.com/datasets/sdot-collisions-all-years/explore?location=47.614507%2C-122.333041%2C12.33
3) City Annual Stats: This dataset includes information of the total population (and change in population), housing, and jobs for the City of Seattle by year. https://data-seattlecitygis.opendata.arcgis.com/datasets/SeattleCityGIS::city-annual-stats-2/explore

Acknowledgements

Khyati Jariwala tidied/joined the original datasets and organized the presentation for our findings. Michelle Xu and Hannah Lee worked on the data wrangling (summary statistics), as well as the data visualizations.