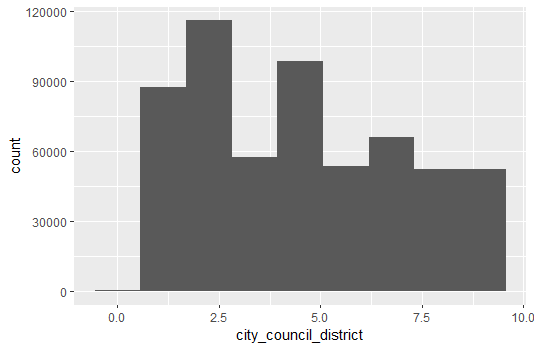
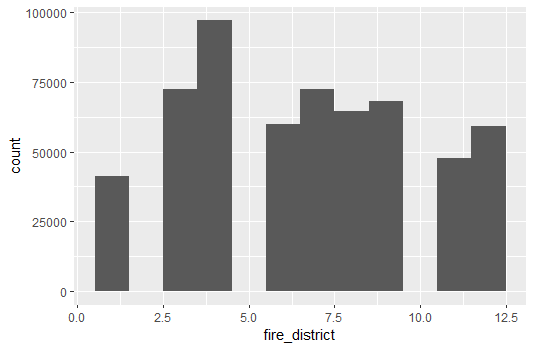
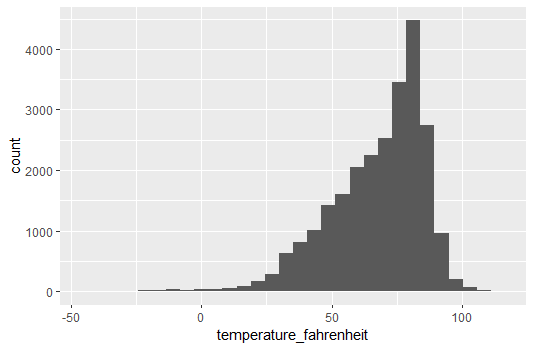
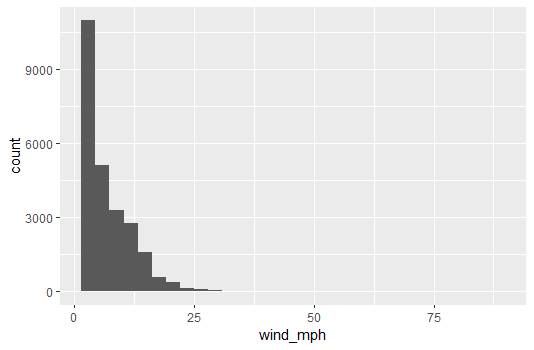
John Incantalupo 1/22/24

CSC-485 Exploratory Data Analysis HW

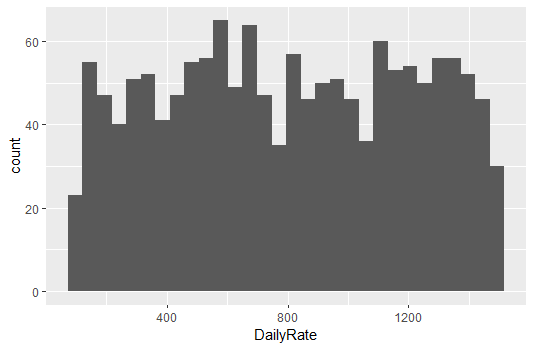
**Boston\_22\_311s.csv / Boston\_23\_311s.csv**

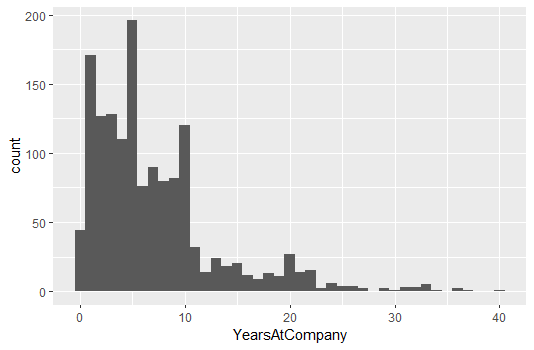
1. Columns: A unique identifier for each observation, the open, closing, and target dates of the case, whether or not it was on time, the status of the case, the reasoning if the case’s status is closed, the title of the case, a few descriptors of the case such as the subject, reason, type, department, address, city district, neighborhood, and latitude/longitude
2. Missing values: There are over 121000 total NAs in the ZIP code column, as well as nearly 4000 NAs in both the latitude and longitude columns. There are also between 5000-7000 NAs in each of the three district columns.
3. Summary Analysis: There are no true numeric variables, so the variables that R recognized as numeric are the district numbers, which is a categorical variable.
4. Basic Plots:
5. Questions: Are there specific types of case that are occur more often in specific districts?

**GlobalWeatherRepository.csv**

1. Columns: Location (including country, coordinate points, time zone), temperature, weather type, wind speed and direction, wind pressure, precipitation, humidity, wind chill, visibility, UV index, air quality based on a few gases, the sunrise and sunset times, the moonrise and moonset, and the moon phase
2. Missing values: There are no missing values!
3. Summary Analysis: There are several data points that produce extreme outliers, such as a time with 87.7 mph winds, or a time with only a 96.8 air quality for carbon monoxide.
4. Basic Plots:
5. Questions: What is the cause of some of the outliers in the data?

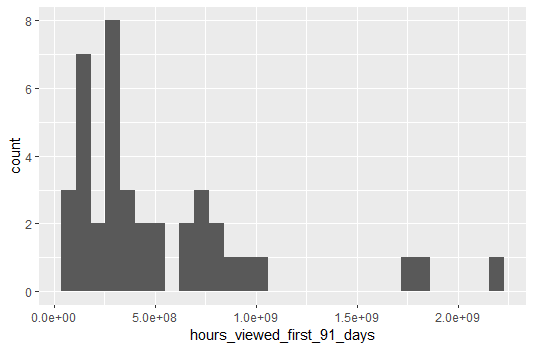
**HR\_Analytics.csv**

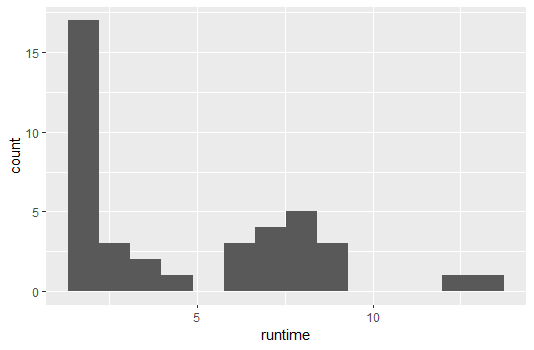
1. Columns: The age of the employee, the daily pay rate, department, distance from home, level of education, a unique employee number, gender, marital status, performance rating, hours per week, years at company and in current role, as well as years since last promotion
2. Missing values: There are no missing values!
3. Summary Analysis: The distribution of the monthly rate is heavily skewed to the right as expected. However, the distribution of each employee’s daily rate is pretty evenly distributed.
4. Basic Plots:



1. Questions: Is there a correlation between the relationship satisfaction and the number of years working at the company? How are variables such as the performance rating determined?

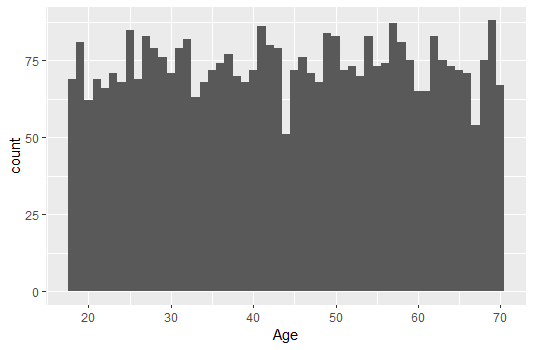
**Netflix-most-popular.csv**

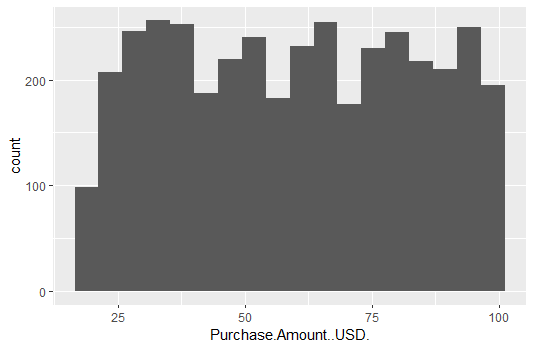
1. Columns: Category, rank within category, title of the show, season #, runtime, and the number of hours and number of views within the first 91 hours
2. Missing values: There are no missing values!
3. Summary Analysis: The English films have, on average, more hours watched than the non-English films. For the most part, this can also be said about the TV shows. Also, there is a pretty strong correlation between the program’s runtime and the number of hours viewed within the first 91 days.
4. Basic Plots:



1. Questions: If we were given data for the hours/views after the first 91 days, which programs would decrease in viewership the most? Are there any program that would maintain their audience?

**shopping\_trends.csv**

1. Columns: A unique identifier for each customer, the name and category of the purchased item, price, size, and color of the item, the subscription status and discount eligibility of the customer, review rating, the number of previous purchases, and method of payment
2. Missing values: There are no missing values!
3. Summary Analysis: Purchases do not exceed $100 but are always at least $20. The lowest review rating given is a 2.5, presumably out of 5. There also does not seem to be a particular age group that makes purchases more or less often than other age groups.
4. Basic Plots:



1. Questions: Are there shopping category that contain more expensive products? Does the size have an effect in the price, as in does the same product have a higher or lower price based on its size?