Nicholas Paisley

Data Visualization

Final Paper

Research Question:

It is easy to speculate that when people have claimed to see a UFO, it can't be true, they are crazy, it is not what it seems, ext. However, what if these people were intoxicated when these claims were made? What if they did not see anything and it was their imagination? Therefore, drunkenness could explain the influx of reports of UFOs.

Describe your dataset:

My dataset consists of 2 csv files:

1. nuforc_events.csv → This dataset has the following categories:
   - i. Event_Time
   - ii. Event_Date
   - iii. Year
   - iv. Month
   - v. Day
   - vi. Hour
   - vii. Minute
   - viii. City
   - ix. State
   - x. Shape
   - x. Duration
   - xi. Summary
   - xii. Event_URL

This dataset had UFO sightings data within it. I used the following Year, Hour, Month and State when exploring the data and Year, Hour and State within the graphs for the data frame. Unfortunately, the dataset did not give whether the UFO sighting was debunked (found out what the object was). This dataset had UFO sightings ranging back from 1061 to as current as 2017.

2. apparent_per_capita_alcohol_consumption_1977_2018.csv → This dataset has the following categories:
   - i. state
   - ii. Year
   - iii. ethanol_beer_gallons_per_capita
   - iv. ethanol_wine_gallons_per_capita
   - v. ethanol_spirit_gallons_per_capita
   - vi. ethanol_all_drinks_gallons_per_capita
   - vii. number_of_beers
   - viii. number_of_glasses_wine
   - ix. number_of_shots_liquor
   - x. number_of_drinks_total

This dataset alcohol consumption per capita data within it. This data, as it says in the title, ranged from 1977 to 2018. In the data exploration, I used number_of_beers, number_of_glasses_wine, number_of_shots_liquor and number_of_drinks_total. Within the dashboard I used number_of_drinks_total.

Data Source:

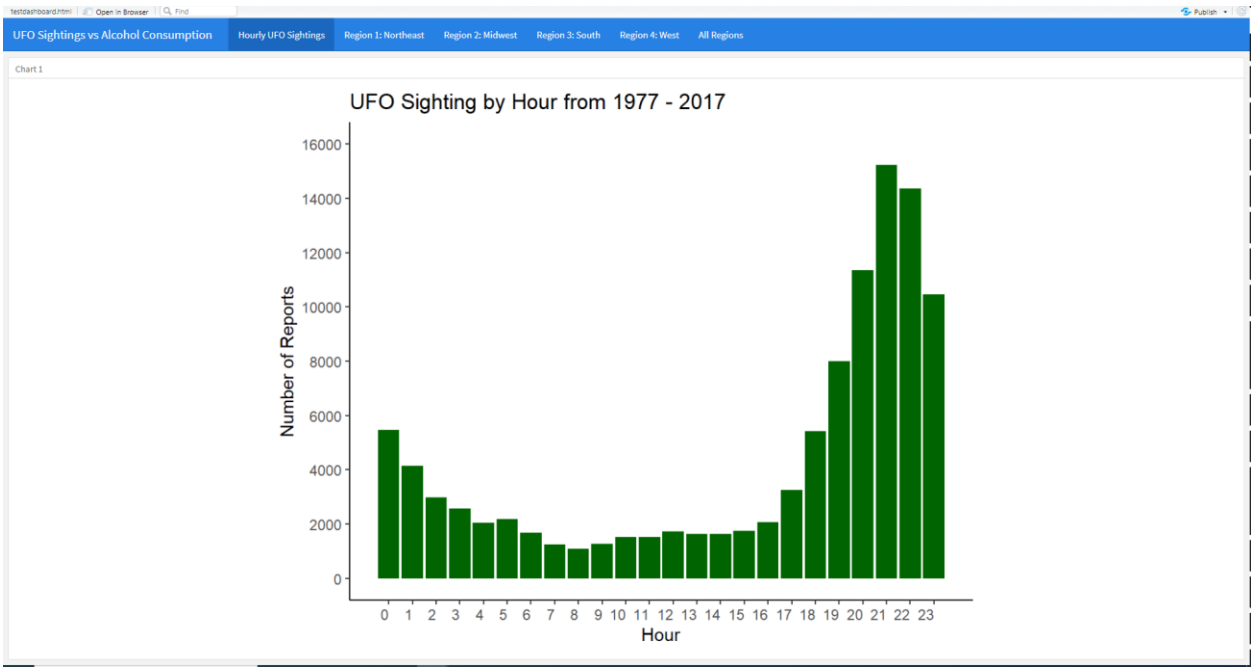
nuforc_events.csv → https://data.world/khturner/national-ufo-reporting-center-reports

This data set was on Kaggle, however, the data was provided by NUFORC (The National UFO Reporting Center (http://www.nuforc.org/)).

apparent_per_capita_alcohol_consumption_1977_2018.csv → https://www.openicpsr.org/openicpsr/project/105583/version/V5/view;jsessionid=DC5A0DD85F252AB9CB0BA2B21D971681

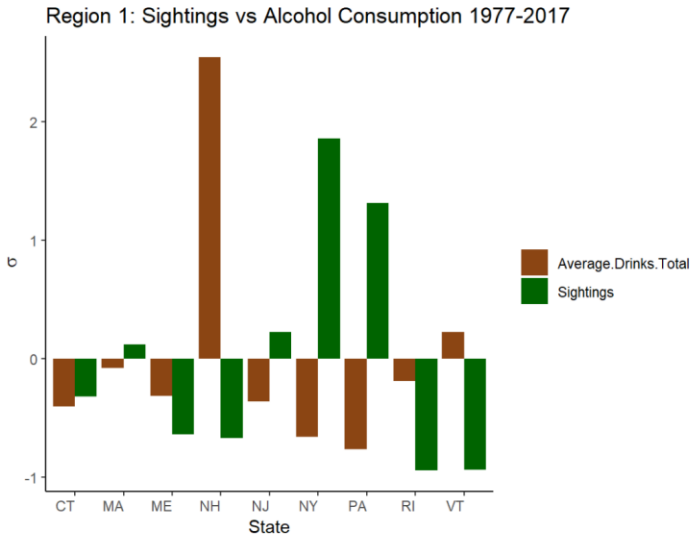
This data set was on openicpsr, however, the data was provided by the National Institute on Alcohol Abuse and Alcoholism.

Dashboard screen shot:



UFO Sightings vs Alcohol Consumption | Hourly UFO Sightings | Region 1: Northeast | Region 2: Midwest | Region 3: South | Region 4: West | All Regions

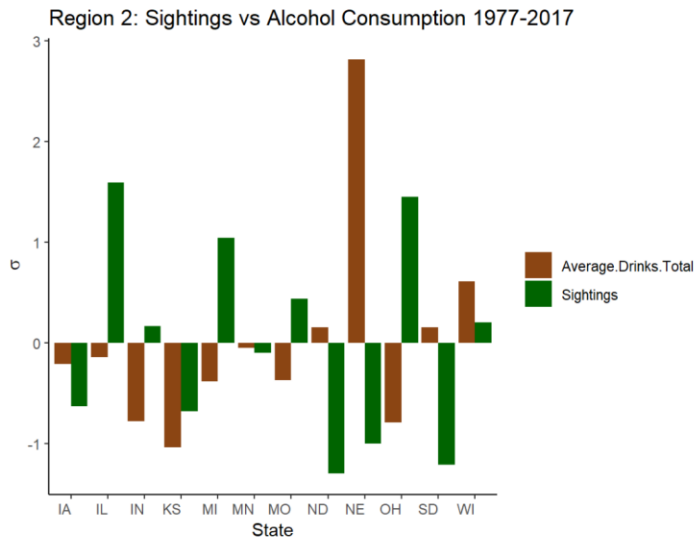
Chart 1



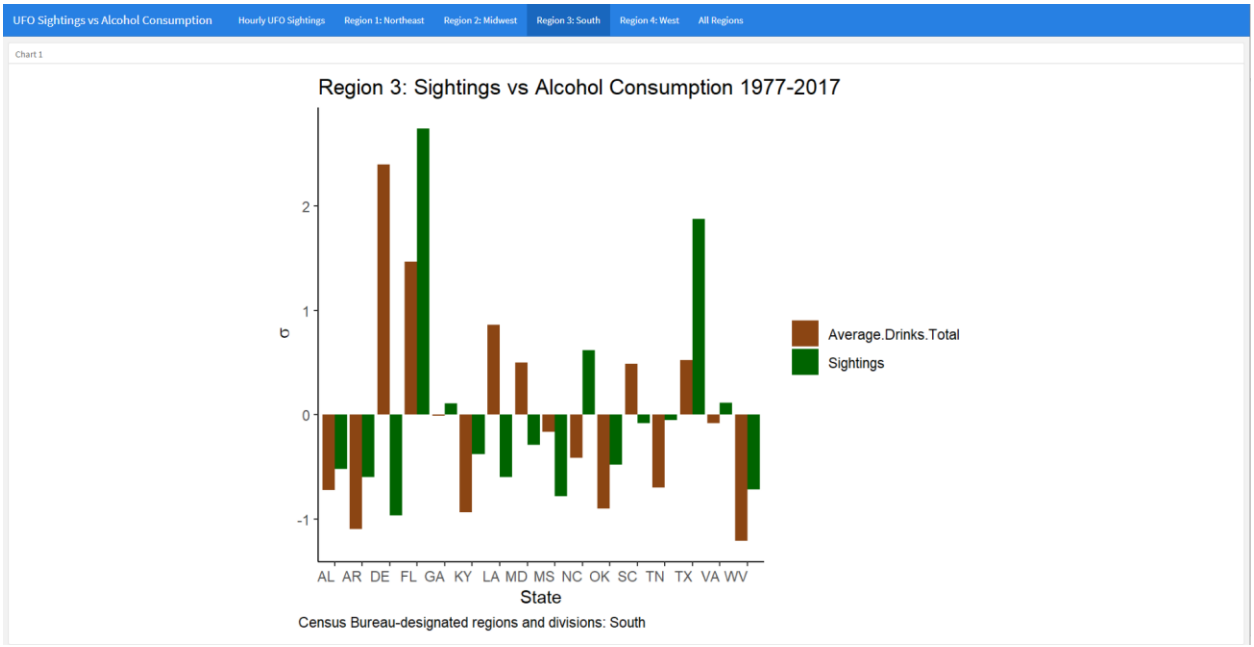
Census Bureau-designated regions and divisions: Northeast

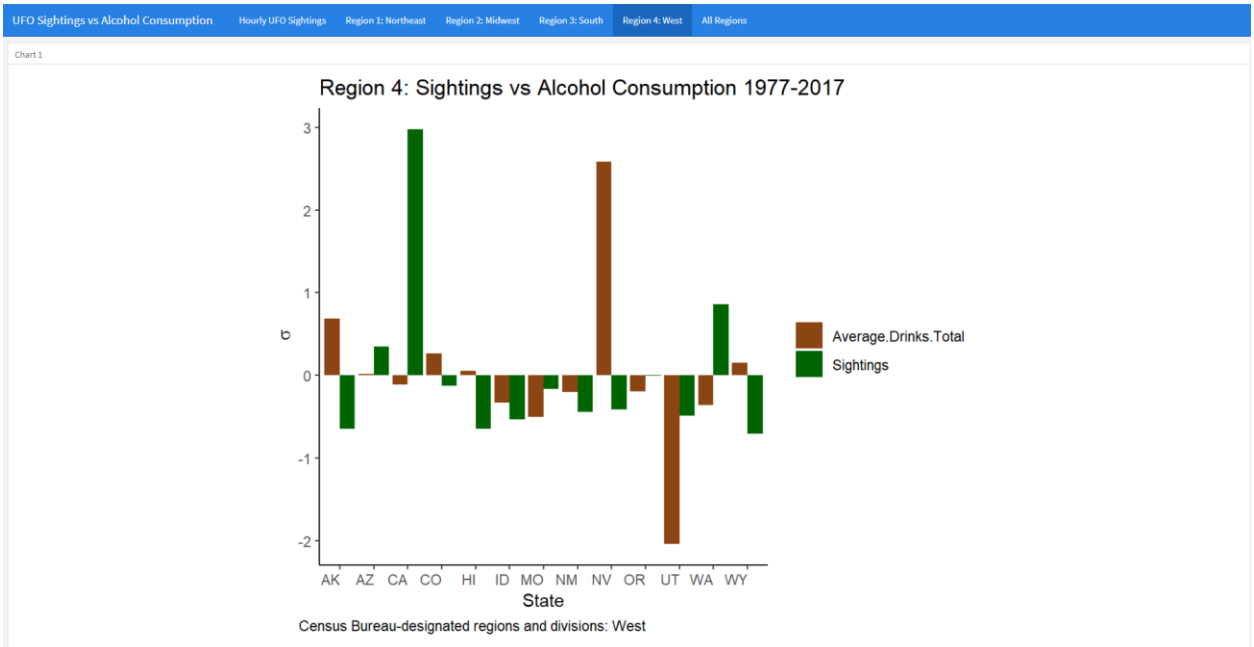
UFO Sightings vs Alcohol Consumption | Hourly UFO Sightings | Region 1: Northeast | Region 2: Midwest | Region 3: South | Region 4: West | All Regions

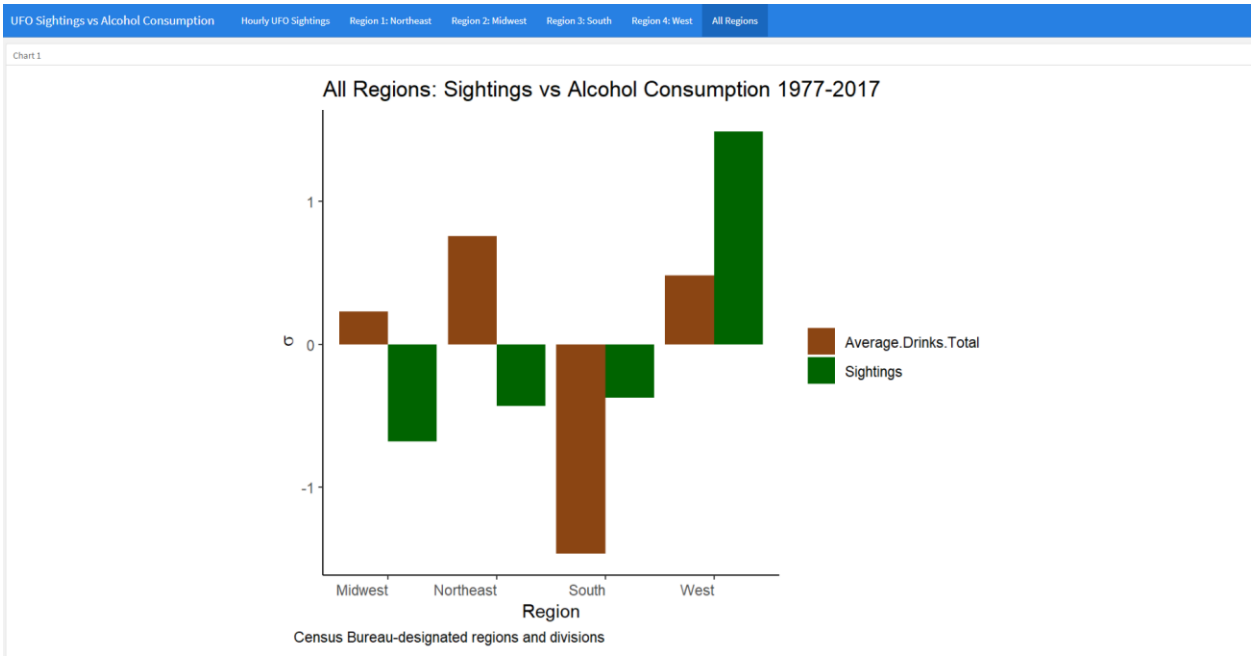
Chart 2



Census Bureau-designated regions and divisions: Midwest


UFO Sightings vs Alcohol Consumption
Hourly UFO Sightings
Region 1: Northeast
Region 2: Midwest
Region 3: South
Region 4: West
All Regions
Chart 1
Region 3: Sightings vs Alcohol Consumption 1977-2017
σ
2
1
0
-1
AL AR DE FL GA KY LA MD MS NC OK SC TN TX VA WV
State
Average.Drinks.Total
Sightings
Census Bureau-designated regions and divisions: South


UFO Sightings vs Alcohol Consumption
Hourly UFO Sightings
Region 1: Northeast
Region 2: Midwest
Region 3: South
Region 4: West
All Regions
Chart 1
Region 4: Sightings vs Alcohol Consumption 1977-2017
σ
3
2
1
0
-1
-2
AK AZ CA CO HI ID MO NM NV OR UT WA WY
State
Average.Drinks.Total
Sightings
Census Bureau-designated regions and divisions: West



Explain the significance of the visuals used in the dashboard to convey the message of the data and the answer to your question.

In order from pages:

1. This graph opens the conversation with the noticing a lot of sightings at night (48.56%)
2. The Regions bring into light whether we see an average between states and alcohol consumption.
    a. Noting the following: There were 4 states that were positive in both sightings and alcohol consumption:
        i. Region 2: WI (Wisconsin)
        ii. Region 3: FL (Florida), TX (Texas)
        iii. Region 4: CA (California)
        iv. All Regions: West
3. The data shows that there is a small correlation that drunkenness could influx the number of reports of UFOs in $\frac{4}{50}$ states were also above the standard deviation in ¼ regions.

*********MORE DEPTH*********

1. **Region 1:**
    a. Drinking
        i. Mean: 576.318 **Standard Deviation:** 149.291
    b. Sightings
        ii. Mean: 1679.22 **Standard Deviation:** 1329.794

2. **Region 2:**
    a. **Drinking**
        i. **Mean**: 550.716 **Standard Deviation:** 144.0898
    b. **Sightings**
        ii. **Mean:** 1564.917 **Standard Deviation:** 1070.917
3. **Region 3:**
    a. **Drinking**
        i. **Mean**: 468.452 **Standard Deviation:** 82.896
    b. **Sightings**
        ii. **Mean:** 1706.812 **Standard Deviation:** 1479.451
4. **Region 4:**
    a. **Drinking**
        i. **Mean**: 563.056 **Standard Deviation:** 128.741
    b. **Sightings**
        ii. **Mean:** 2572.231 **Standard Deviation:** 3233.308
5. **All Regions:**
    a. **Drinking**
        i. **Mean**: 539.636 **Standard Deviation:** 48.594
    b. **Sightings**
        ii. **Mean:** 1880.796 **Standard Deviation:** 465.032

Explain the data visualization concepts.

Kosslyn:

1. **The Principle of Relevance:** I used this principle in all my graphs. I wanted to show the effectiveness of my points with as little information to not overwhelm my audience, where they can see my point very clearly and objectively through all of my illustrations.

2. **The Principle of Appropriate Knowledge:** I am assuming that the audience understands military time in this graph. With 0 being 12:00 am (midnight) to 23 (11:00pm). I am also assuming that everyone in my audience knows the abbreviation for Unidentified Flying Object (UFO). This is also true for the abbreviations of the states within the region graphs.
3. **The Principle of Salience:** We are able to see a clear difference in the graphs. When referring to the "UFO Sightings by Hour from 1977-2017" graph, we are able to clearly see a change within the data. The same thing for the "Region" graphs. They are standard deviation graphs, but you can clearly see the changes within the observations.

Tufte's:

1. **The representation of numbers, as physically measured on the surface of the graph itself, is directly proportional to the to the numerical quantities represented.** This is only for the "UFO Sightings by Hour from 1977-2017". The graph is showing some the measurements hitting over

10000 observations. This makes sense because we are surveying data over 40 years. It is justifiable that these numbers are not out of proportion for that time span.

2. **Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graph itself. Label important events in the data.** The labelling in the graphs is accurate and precise. The labels and title clearly state what the graph is showing and representing. For the title, it states what we are looking at "UFO Sightings by Hour" and the time span that is being observed "from 1977 – 2017". Also, "Hour" on the x-axis clearly shows all the individual hours. Again, the hours are in military time. The "Number of Reports" on the y-axis depicts how many reports were made over the time span. Also, on the y-axis, the intervals were made to show every 2000 to help the audience be able to get a better representation on how many observations were observed. When it comes to the other graphs, the "Regions" graphs, it is obvious that the x-axis, state, shows the state abbreviations. The harder one for explanation is the y-axis, the "**σ**" label. This "**σ**" is actual the mean of the standard deviation, it signifies what the how much the data varies from the average.
3. **Show data variation, not design variation.** I am clearly showing the data is different with the peaks of data. There are some that are close to similar, however, they are not the main focus of the graph (the max and min 'hour' are what is going to grab the attention of the audience). This also the case with the standard deviation (**σ).** You can how the data varies from the average in the graphs, small or large.

Berinato:

1. **Data-Driven and Declarative (Everyday dataviz):** The graph is representative as a "Everyday dataviz". This falls in this form because it is a simple graph (bar chart) and it is *simple*. It comes with the factual information and based on the data. Kept the color non-stigmatizing to try to not pull the audiences view away from the point of the graph.
2. **When a chart hits your eyes principle, "We see first what stands out":** The graph will show the main focus of the max and min with the associated time associated with it. In this case we are looking at "UFO Sightings by Hour", and we can clearly see when the peak UFO Sightings are apparent. When it comes to the region graphs, it is more on the focus on highs and lows.
3. **To make charts that look elegant or beautiful, focus on design *simplicity*:** I am showing only what I need within and only the colors I need. I tried to limit the amount of eye travel and not over confuse the labels.