Kevin Kalunian

Professor Michael Salé

Business 460

Comcast Descriptive analysis-R Project

01/19/18

**Comcast: Customer Service, Value of Service, and Technical Services are Rated from Zero-to-5 Stars in a Revealing Case Study Measuring Complaints from 45 States and the District of Columbia**

Background

Comcast is a producer of feature films and television programs intended for theatrical showing and over-the-air and cable television broadcast. Comcast was criticized for a multitude of questionable business practices and extraordinary customer dissatisfaction, and was positioned behind other industry peers from the cable industry. Comcast Corporation, an American global telecommunication giant, dominates in broadcasting, and cable television around the world. It is the second-largest pay-television company and largest home Internet service provider in the United States, and the nation's third-largest home telephone service provider. Comcast provides services in the U.S. residential and commercial areas in 45 states and in the District of Columbia.

What were the customers major complaints when they called Comcast's customer service centers? Were those issues chiefly technical-bad installation or equipment, quality of products and services (value oriented), or based on a bad experience with Comcast’s customer service personnel? Was one region mistreated more, based on the U.S. Census geographic breakdown by region? Which region was more outspoken with their complaints than the other regions?

Main Focus on Study

Are customers cancelling service because of technical issues, billing issues, of valuation of services? What type of analytical tools were used?

I selected descriptive analytics, getting an insight into the past and summarized raw data, making my report consumable by everyday consumers. The dataset’s text remarks described the past events. April 8, 2008 through September 9, 2016 was the timeframe that these events occurred. Descriptive analytics are useful because it allowed me to learn from the customer’s past behaviors, and to understand how the Comcast might benefit in the future from these outcomes.

Choosing the Project

Most consumers can relate to the common problems with technology services and products like cable television. I am a customer and shareholder of Comcast. I also worked as a freelance technician for the NBC News Bureaus, and had a vested interest in the story and outcome from the eight-year study. I also thought the results and comments would be useful as an investor, and provide additional entertainment by reading these complaints.

I used JMP to create reports that provided historical context to the company’s problems and their and customers’ experiences. The descriptive statistics allowed me to understand what the data showed. Reading through the data set, (text) and understanding the basics of the study gave me a greater appreciation of what data scientists do and how they must overcome shortfalls in the data as well as the errors within the text, and errors that they can contribute via harried workflows, tainting the report.

The Tools and Processes Used

Organizing and optimizing the data set in the JMP application.

Eliminate superfluous columns, combine columns.

Use delimiter to remove words in column to create separate columns for author's text, city, and state for segregation, and further analysis.

Analyze > Text Explorer >

Sort and filter consumer reviews with a star rating system. Repeatedly read the text remarks to verify that 0 is the worst and 5 is the best on the scale system.

Cleanse and analyze a random sampling of reviews. Repeat the process after errors were committed.

Determine grouping percentages and breakdown reviews by 0, 1, 2, 3, 4, or a 5-star rating system.

Create geographic footprint for data based upon regions.

Use an aggressive approach for text cleansing due to the texts' authors inability or harried writing of remarks or other data, even though said remarks were mostly civil. Some used mild profanity, but from over 5,500 complaints, no “F-bombs” were used, unless they were previously cleansed.

Errors in aggressive data cleansing were corrected a third time, operator error occurred.

Writing code in R Studio to filter states into regions for geographic analysis.

A Conclusion

Converted raw code ratings into 3 groups for coarse review, i.e. the 2, 3, 4, or a 5-star rating accounted for only **6.5%** of all respondents, **93.5%** of Comcast's customers from all regions rated Comcast's services, (*technical, costs, or customer service*) negatively with a 0 or 1 star rating. Only 32 respondents failed to enter a text remark. I conducted an analysis state-by-state looking for customer issues with Comcast and considered the market share, learning about different providers and the states and the regions they are locations in the United States. New York, located in the Northeast region, had fewer customer complaints because Time Warner Cable is the state’s main ISP. Other regions had COX, CenturyLink, Midcontinent, Charter, Mediacom, and Frontier as those state’s utilities. The dataset only evaluated the customers that used Comcast’s services. Some states were broken up fairly evenly between ISP, others were dominated by one company and it was not within the scope of this report to make judgement calls about monopolies, the customers spoke to that issue over three-hundred times in this study. According to an article in Recode.net, April 27, 2017, Comcast has 25.1 million broadband subscribers.

As a result of Comcast’s poor relationships with their customers, the FCC hit Comcast with the biggest cable fine ever, according to Money.cnn.com’s web page from October 12, 2016.

Additional Notes of Interest

Raw complaint data about Comcast television and the internet was published at consumeraffairs.com.

New York State only had 65 respondents and this seemed to slant the study away from the Northeast region. Time-Warner Cable is the main provider for New York. Alaska and Nebraska both had only one respondent. Iowa, Montana, Oklahoma, Rhode Island, and South Dakota was not included because I assume those states have other exclusive providers when the study was initiated. I’m unsure of the pattern leading up to cancellation of services. Comcast’s customers in general felt that they were being “*ripped-off*,” and 22 mentioned they would *call or write* to the Attorney General in their home state.

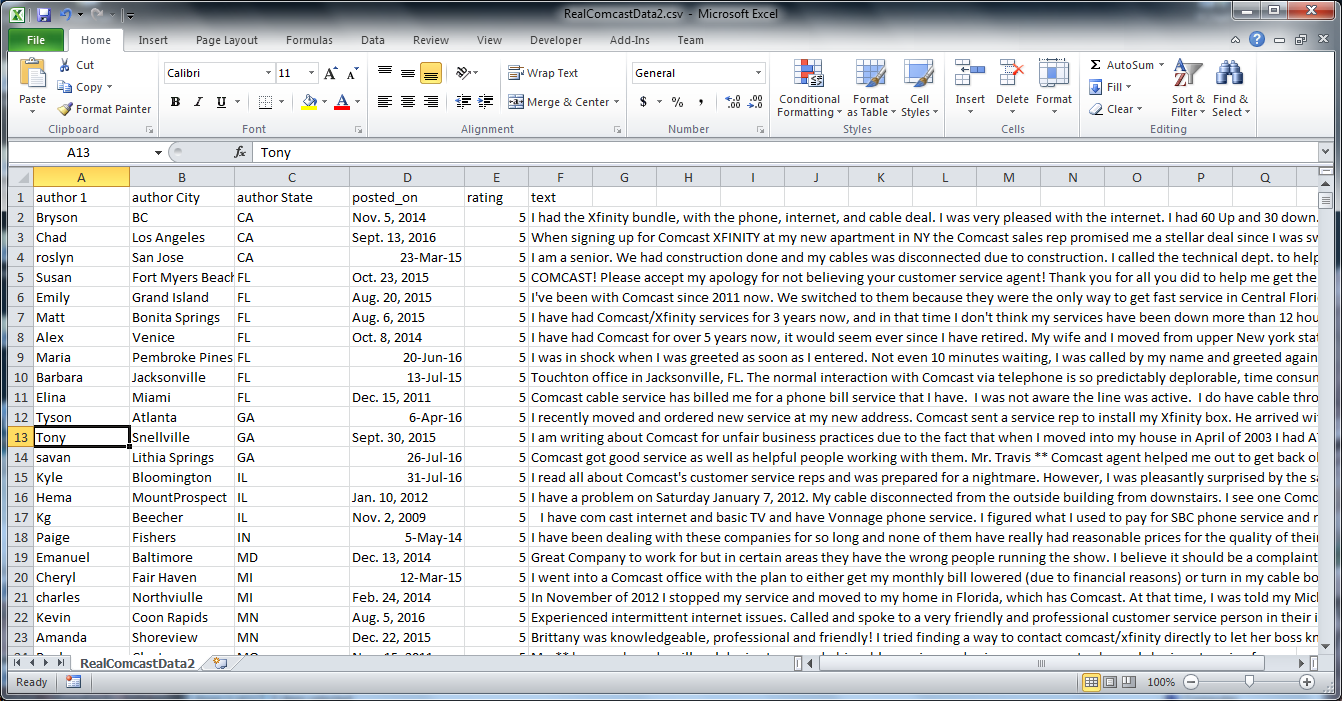
How It Worked or Didn’t Work as Planned

As you can see, I have computed that most of the complaints seem to come from the South. Interesting to note also is that 23 entries didn’t fit into any portion of the US. This means that there were most likely some data entries with bad State data. More data cleaning was required in JMP. Through this project, I learned quite a bit about JMP and R. My programming experience in Intro to Java made programming in R only slightly easier.

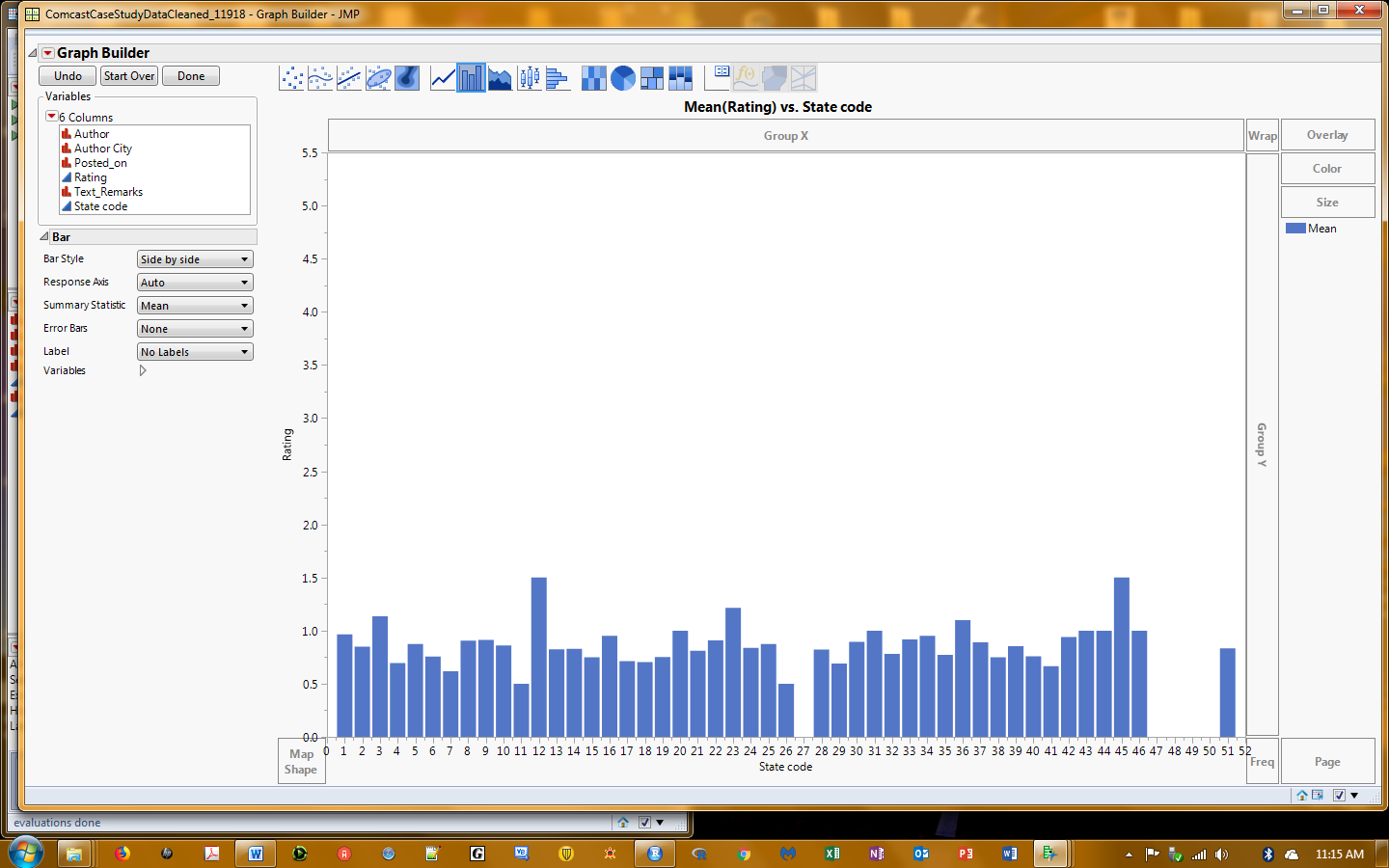
As part of the final project, I wanted to use JMP and R to make some conclusions about the Comcast data set. I did quite a bit of the R online tutorials, and I was surprised to find that I was able to put some of that knowledge to use for the final project. Combining JMP and R, I was able to analyze and draw some conclusions about the Comcast complaint data.

The original data set contained a column containing Person Name “of” City, State, then a column with a date, a rating, and then a column with a long descriptive text describing the problem/complaint. In order to make the data useful in R, in JMP, I replaced “of” with a “:”. I was then able to split that first column into three separate columns: Name, Town, and State, keying off the “:” and “,” as delimiters. This was fairly straightforward to do in JMP.

I then spent some time cleaning the data in JMP. All the states were converted to two capital letters. With this done, I exported the data as a csv file. A sample is shown below.



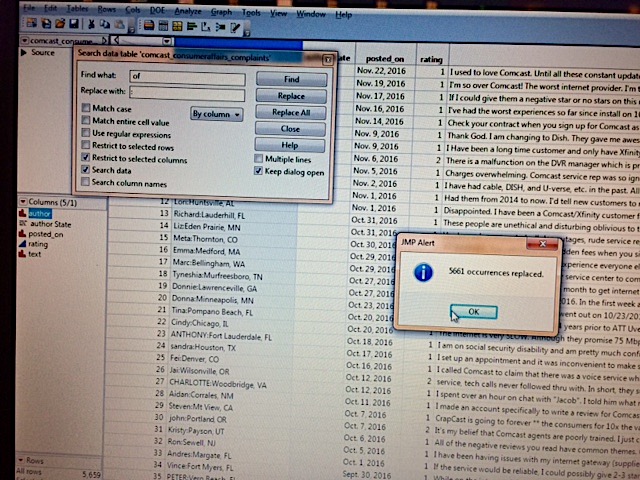
I was then able to read in this CSV file in

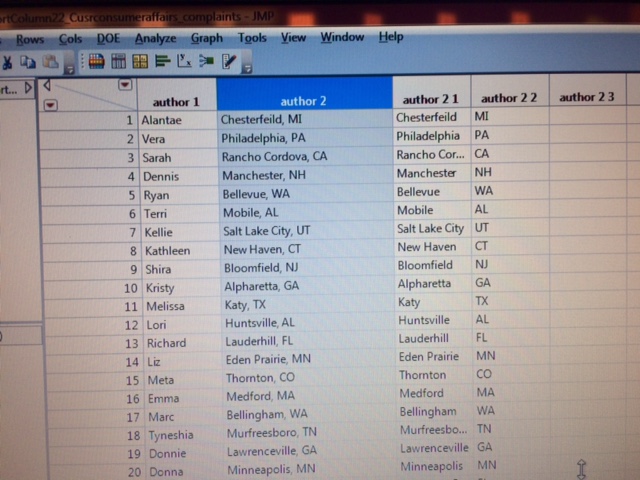


JMP Graph with some insufficient data such as state number 27, which had only one respondent does not show up in the graph…state number 51 are the 5 outliers without a state identified.

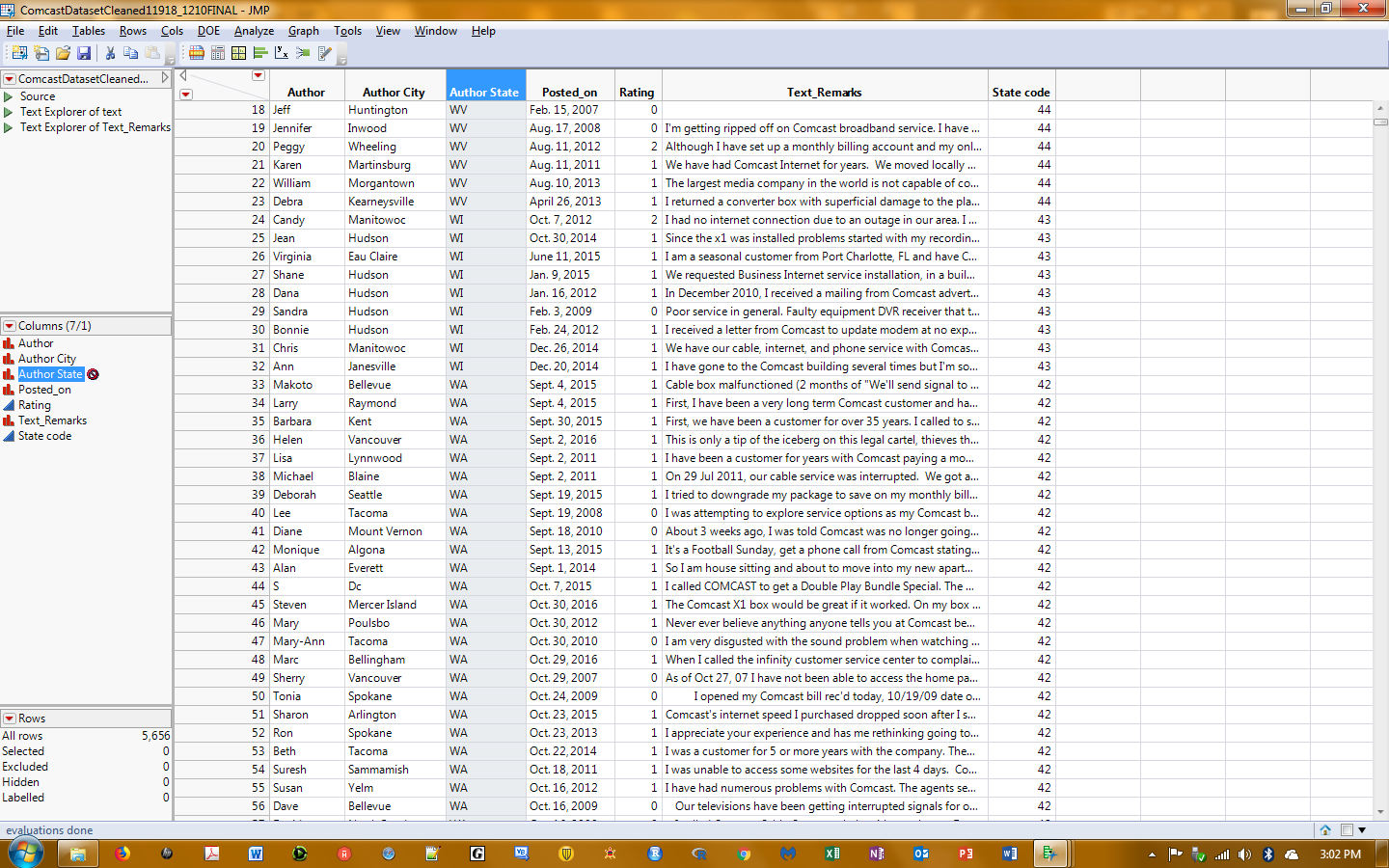
Some jpegs of the labor-

Begin a delimiter process to break apart this string of author’s name, city, and state.

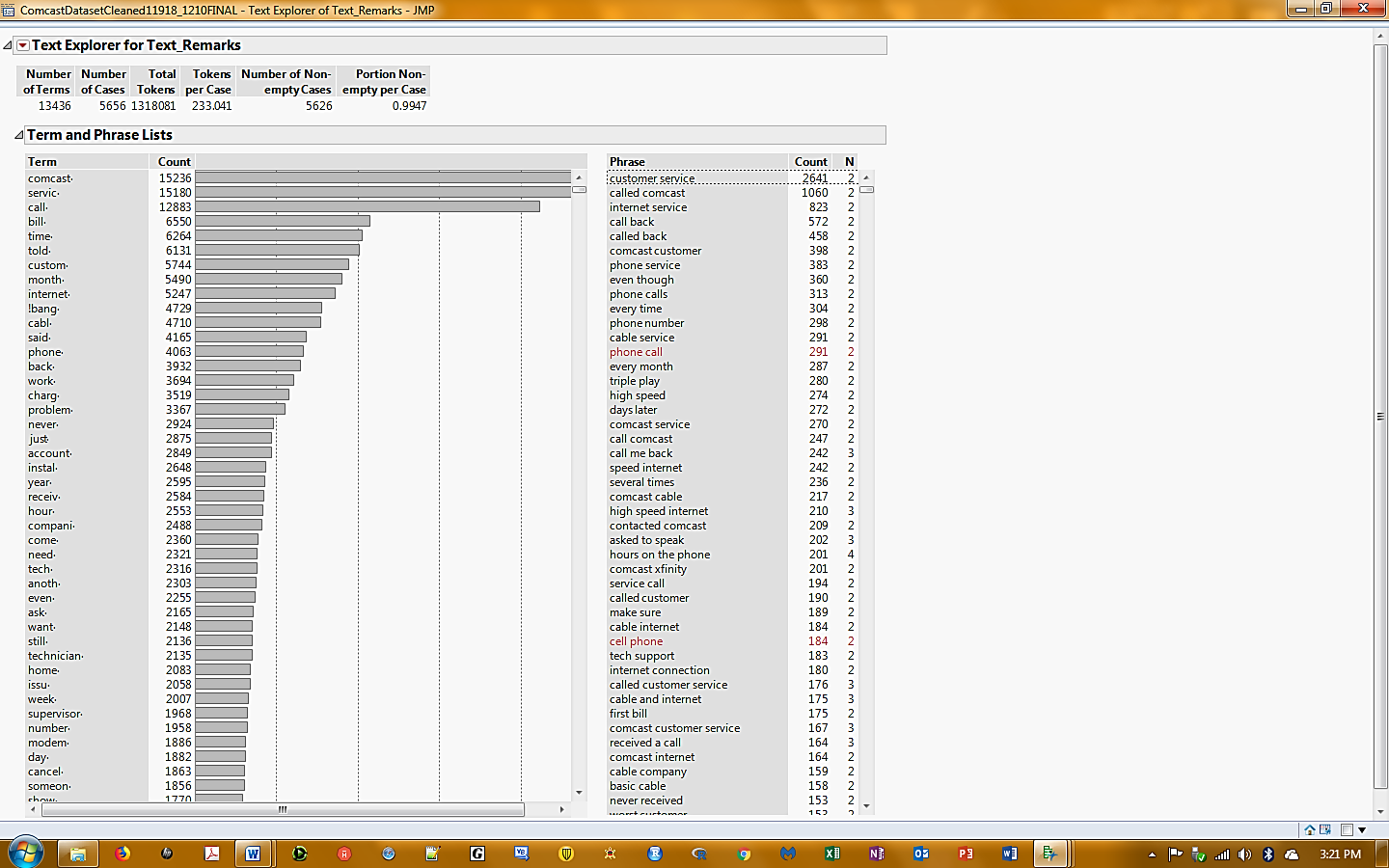




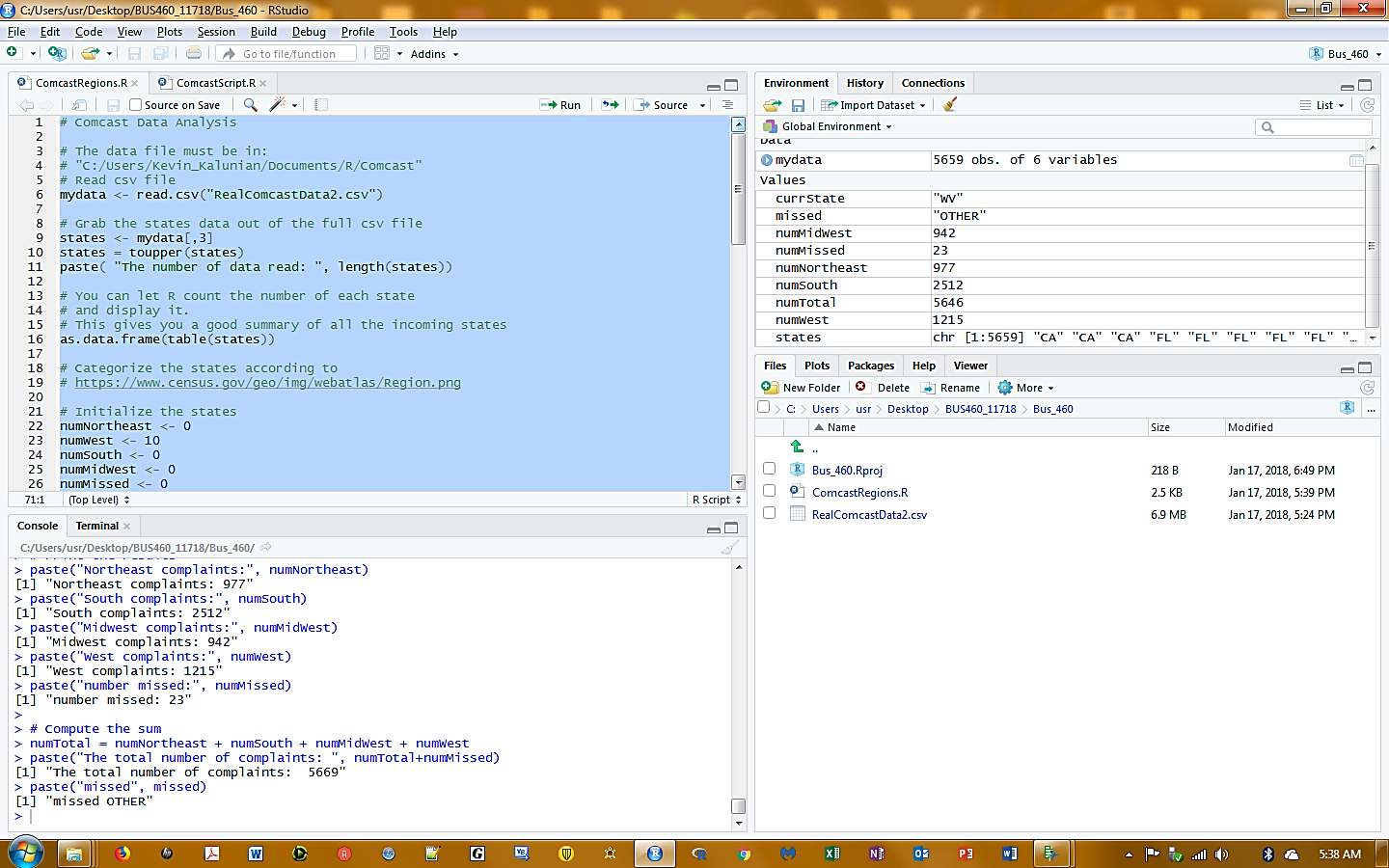
The “Cleansed” data set in JMP.



“Bag of Words” sorted to a few tokens by frequency, and a two-to-four word phrase using Text Explorer.



My project workspace in R Studio looked like this:



The program I wrote uses data reads, data tables, a for loop, and an if statement. I made numerous errors, then corrected those. These are all topics that I learned in the R tutorials of which I did not complete before the deadline.

Here’s the Required Output for the Project

> # Print the results

> paste("Northeast complaints:", numNortheast)

[1] "Northeast complaints: 979"

> paste("South complaints:", numSouth)

[1] "South complaints: 2517"

> paste("Midwest complaints:", numMidWest)

[1] "Midwest complaints: 944"

> paste("West complaints:", numWest)

[1] "West complaints: 1210"

> paste("number missed:", numMissed)

[1] "number missed: 6"

>

> # Compute the sum

[1] "The total number of complaints:  5656"

> paste("missed", missed)

[1] "missed NEW ZEALAND"

>

> # Display the complaint breakdown from “Bag of Words” then “Tokenizing” them

> paste("The number of Tech complaints", numTech[1], numTech[2], numTech [3], numTech[4])

[1] "The number of Tech complaints 724 1959 727 929"

> paste("The number of Billing complaints", numBilling[1], numBilling[2], numBilling [3], numBilling[4])

[1] "The number of Billing complaints 148 337 125 169"

> paste("The number of Value complaints", numValue[1], numValue[2], numValue [3], numValue[4])

[1] "The number of Value complaints 40 79 27 24"

TOKENS from “Bag of Words”

Technical issues

PROBLEM, CABLE, INTERNET, MODEM, EQUIP, INSTALLATION

Billing issues

CHARGE, PAY, BILL

Value of Products and/or Services

CONTRACT, HORRIBLE, CHARGES, CUSTOMER SERVICE, ATTORNEY GENERAL

THE OUTCOME

|  |  |  |  |
| --- | --- | --- | --- |
| Comcast Customer Complaints - Breakdown by Regions and Categories via R program | | | |
| U.S.  Region | Technical Problems | Billing Complaints | Value of Products or Services Complaints |
| Northeast | 724 | 148 | 40 |
| South | 1959 | 337 | 79 |
| Midwest | 727 | 125 | 27 |
| West | 929 | 169 | 24 |

**Northeast**

MA, ME, VT, NH, CT, NY, PA, NJ

**South**

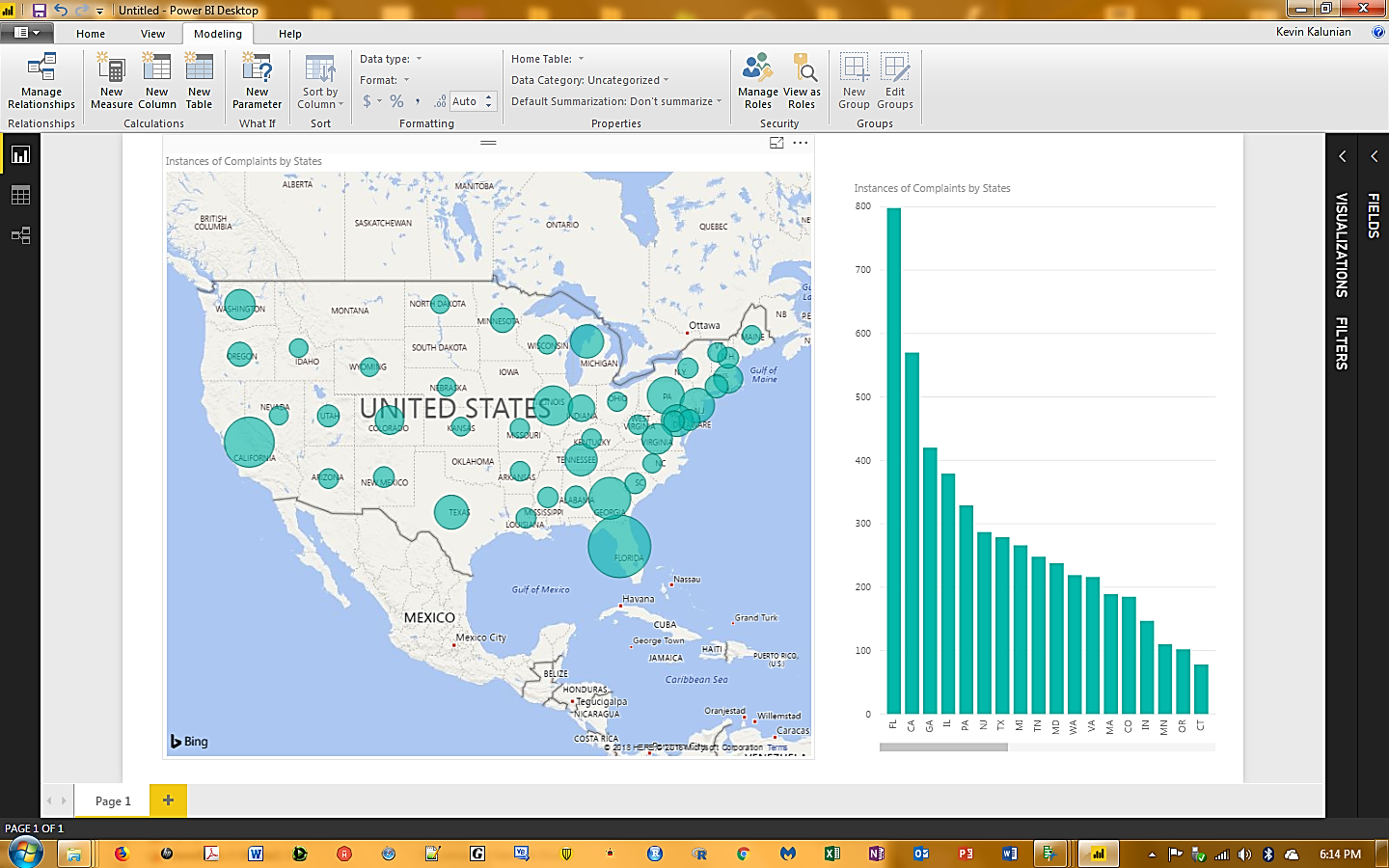
TX, AR, LA, MS, TN, KY, WV, VA, NC, MD, DE, AL, GA, SC, FL, DC

**Midwest**

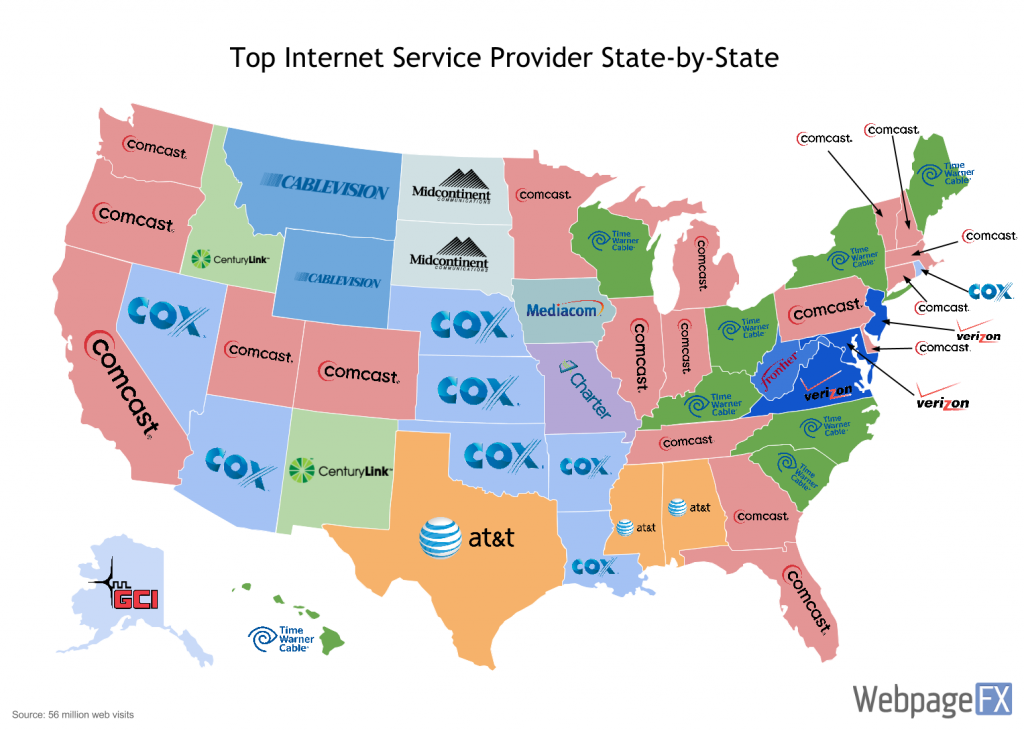
ND, MN, WI, MI, NE, IL, IN, OH, KS, MO

**West**

WA, OR, ID, WY, CA, NV, UT, CO, AZ, NM, HI, AK



Power BI Output with map and data from each state and the bar graph with same data.



Comcast in represented in 45 States and the District of Columbia.

<https://www.webpagefx.com/blog/internet/who-controls-the-internet-a-state-by-state-look/>

Some references:

KAGGLE.COM URL link to .csv file

<https://www.kaggle.com/archaeocharlie/comcastcomplaints/data>

<http://money.cnn.com/2016/10/11/news/companies/comcast-fine-fcc/>

https://www.census.gov/geo/img/webatlas/Region.png