**Introduction**

I spent my undergraduate years studying European history, which emphasized the intersection of many diverse disciplines including: politics, warfare, anthropology, economics and sociology. As a U.S. Army Reserve Officer’s Training Corps Cadet with a future as an Army officer, my intent was to gain insights into the decisions that populations, institutions and governments make in order to prepare me to operate in uncertain environments. This understanding was valuable to me as a junior officer because it allowed me to empathize with people from disparate cultures, understand their plight and connect with them on a personal level to support our missions. The majority of my work was conducted in the field, with little more than a pen and paper and it was often chaotic and fluid, requiring our team to react to situations on the ground and make hasty decisions that would have far-reaching secondary and tertiary effects.

I gained additional rank and responsibility over the next few years and found myself joining the U.S. Army Civil Affairs community, which is an organization under the United States Army Special Operations Command. The Charter of the civil affairs branch is to collaborate with stakeholders ranging from the U.S. Department of State, United States Agency for International Development and various intergovernmental organizations and nongovernmental organizations; to apply a multifaceted approach to assist the public health, education and humanitarian assistance needs of partner nations across the world. From 2015 through 2018, I put the skills I learned into practice, overseeing operations in Southeast Asia, and leading a development team through the steppes and deserts of Mongolia.  As the U.S. Ambassador to Mongolia’s primary infrastructure and disaster response advisor, I had undertaken a profession that demands excellence in planning, executing, and measuring the efficacy of multi-million-dollar development programs.

What I learned over these years is that my humanities background and the training I received in the U.S Military, while beneficial did not adequately prepare me to perform my duties to the best of my ability. My training was centered on qualitative methods, which included case studies, anecdotal evidence and biased first and second-hand accounts from those with whom I interacted. There was very little focus on quantitative approaches, namely that our data collection methods were not standardized, we did not use statistical techniques to perform tests and did not hunt for patterns in our data to generate new insights. Just as I had realized that a multi-disciplinary approach was requisite to understanding history, so too did I believe that combining new methods of analysis would facilitate my bosses’ visualization of the operational environment and support their decision-making processes.

In seeking higher education, I realized that Applied Data Science would be the ideal field of study for this task as it would revolutionize how I collected, prepared, analyzed, managed and preserved data, and most importantly, how I would communicate recommendations based on my findings in the data. What I learned from the ADS program of study was that my previous approaches to problem solving were unstructured and disorganized; I did not consistently see the larger picture or effectively seek stakeholder and domain expert advice, to the detriment of my work. What I strive to show over the next several projects is how the studies I completed at Syracuse University’s ISchool changed my analytical approach, improved my communications skills and has further prepared me to take positions of increasing responsibility as an operations director overseeing regional aid and disaster assistance programs.

I will start with examples from my Introduction to Data Science Course, which I particularly appreciated because it brought together new students to data science from many different fields. In this class, I gained high-level understanding of Applied Data Science concepts, and I was fortunate enough to work with an operations director from Microsoft, a self-employed financial analyst and veteran programmer, three distinctly different fields. After this, I will illustrate how my Database Management Courses altered how I approached data collection and management, bringing to light new approaches that I had never considered. Lastly, I will use my Data Analysis and Decision Making classwork to demonstrate how gaining mastery of hypothesis testing and iterative modifications to processes will support future humanitarian assistance programs.

**Case Study 1**

*Introduction*

This first project for IST 687, Introduction to Data Science was an important primer for my Applied Data Science studies. Here we were taught not only the basics of Data Science, but how to apply what we learned to a relevant, real-world situation. In my previous field work, I was accustomed to gathering data from multiple sources, including first-hand accounts, other organization’s reporting and open-source material, but did not know how to combine that data to produce meaningful analysis or isolate the driving factors of our target variable. For example: in Mongolia in 2017, my team was conducting research to determine locations best suited to develop education infrastructure. Our goal was to reduce poverty and increase education opportunities for young children. To inform our recommendations to the Defense Security Cooperation Agency (DSCA), we conducted site surveys, population assessments, open-source research and considered Mongolian government officials’ inputs. Issues arose as we received conflicting reports from all angles. Stakeholders had vested interests in allocating capital for projects that would grant them political credibility or personal gain, rather than support the most destitute populations. We had to significantly delay or cancel projects when we could not identify which projects would serve the most deserving communities. The stakeholders at DSCA were not comfortable approving multi-million-dollar projects unless we could support our claims with hard evidence. Had we a possessed a foundational knowledge of Data Science concepts, we would have had a better idea of which data to collect and how to apply the proper models to make more effective decisions. This first project illustrates those concepts.

In this project, our goal was to analyze the housing market around Sacramento, California and make purchasing recommendation to the client. We sought to answer the following questions— “What is overall price range for the market? What is the average price for a five-bedroom home? Which property feature is the best measure of cost? Will a $450,000 budget afford us the ability to purchase an additional, profitable rental property?” Our platform of choice was “R” because the dataset was relatively clean already and we could apply existing functions quickly.

*Methods*

We obtained data from the Sacramento Bee newspaper website which was formatted in a CSV file that spanned a 7-day period and included 985 real estate transactions in that region. After cleaning the data, we settled on 12 variables, including things such as “square feet”, “beds”, “baths”, “city”, and “zip code”.

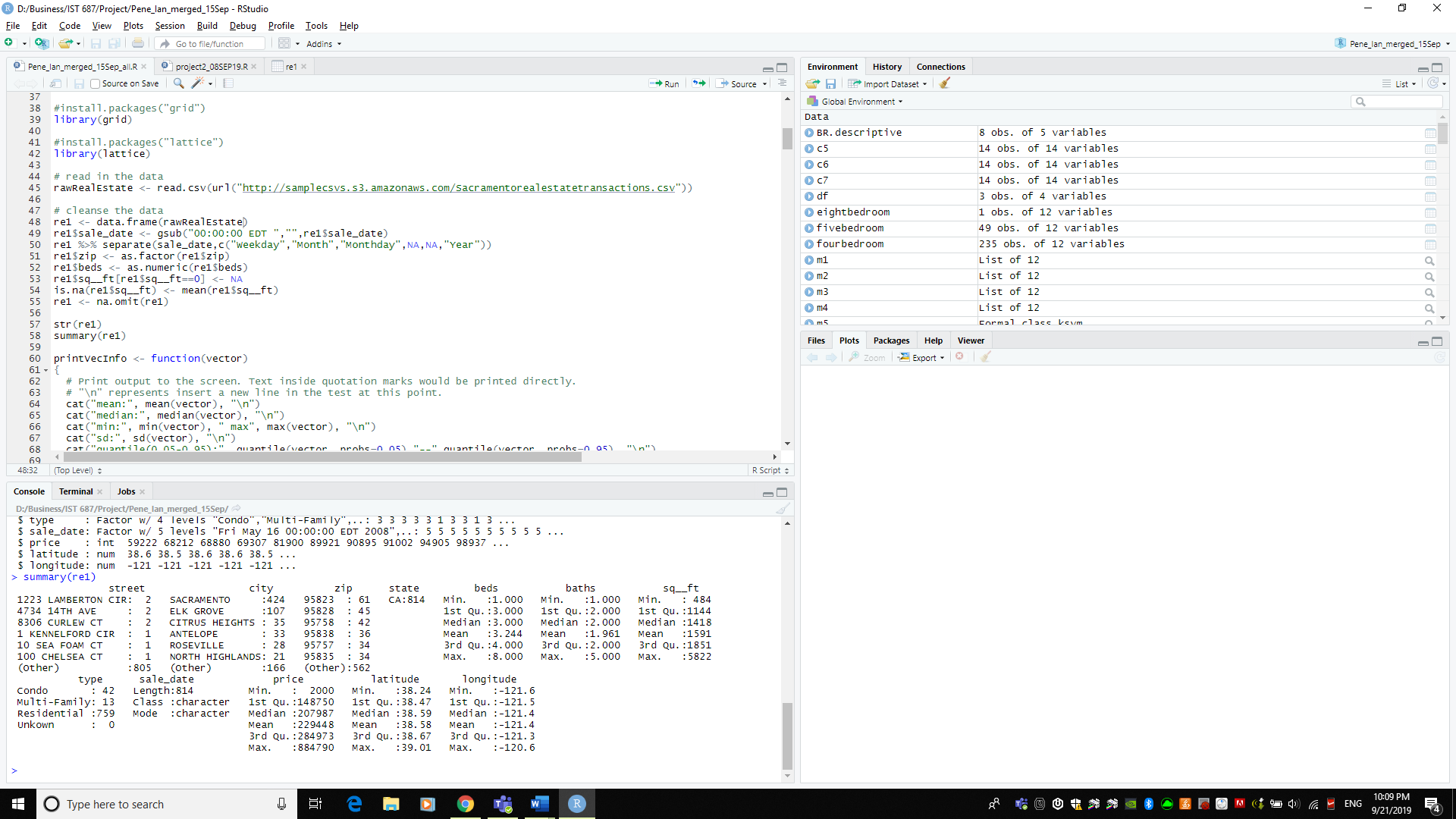


Fig 1) Summary of Raw Housing Data

Our intent was to conduct exploratory data analysis and summary statistics to identify any unusual or interesting patterns in the data. One of the first interesting things we encountered was that multifamily homes with 6 bedrooms often cost much less than single family “residential” homes.

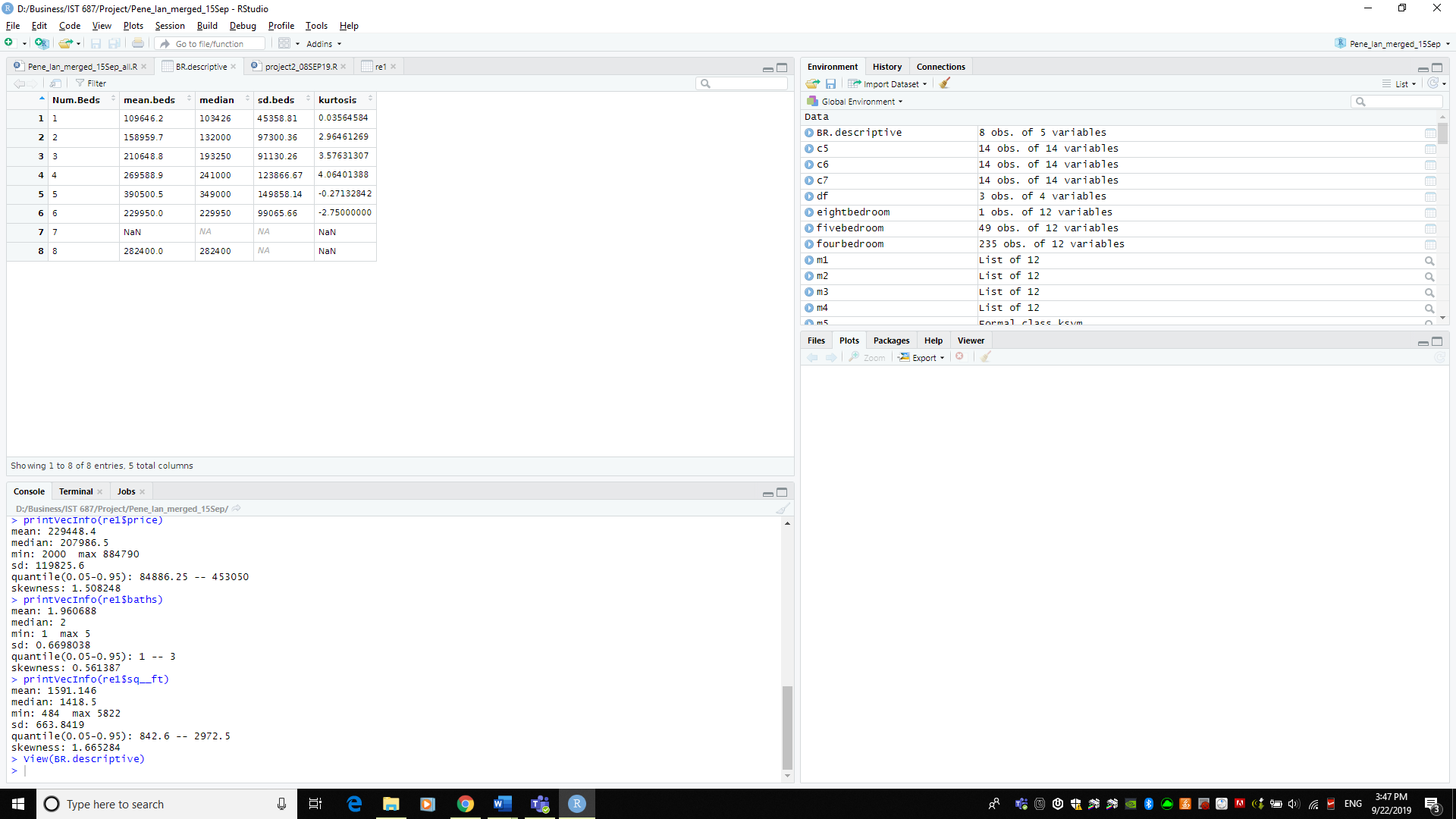
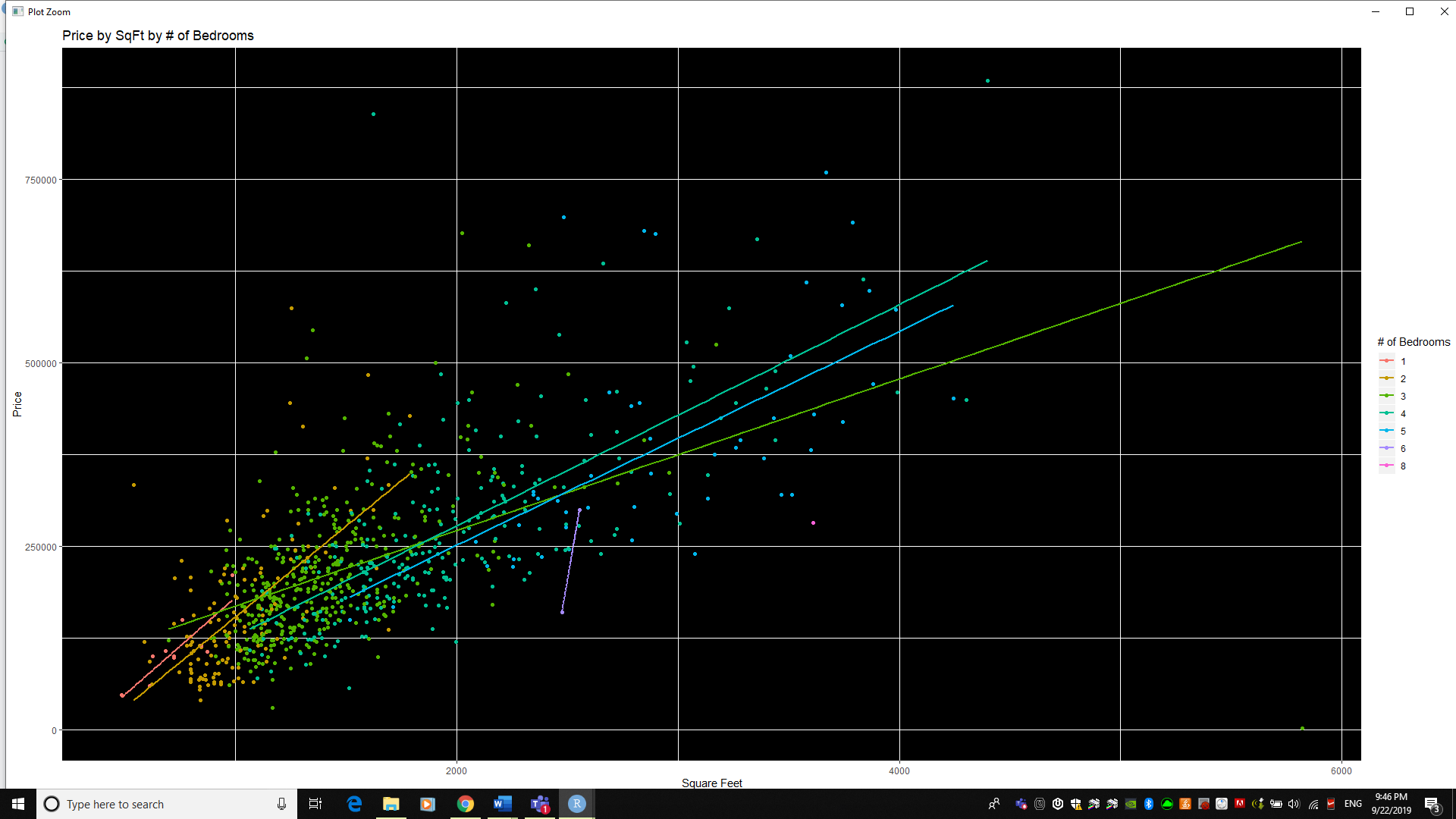


Fig 2) Bedrooms Not Only Factor in Price

This is where I believe the significance of data science concepts first shined for me. Often, we are presented with evidence that implies a causal relationship, and we may incorrectly latch onto the first justification to make our recommendations. In this case, we delved deeper into the data and continued to identify patterns which more clearly explained housing price, namely “lot size”, “square feet” and “zip code”.

Fig 3) Beds / SQFT More Accurately Predict Price

In this course, we learned the value of combining datasets and using more than one approach to reach our recommendations and conclusion. After our exploratory data analysis, we decided to overlay our data set on a map of the region by scaling a KMZ file to match the longitudinal and latitudinal values assigned to each house.

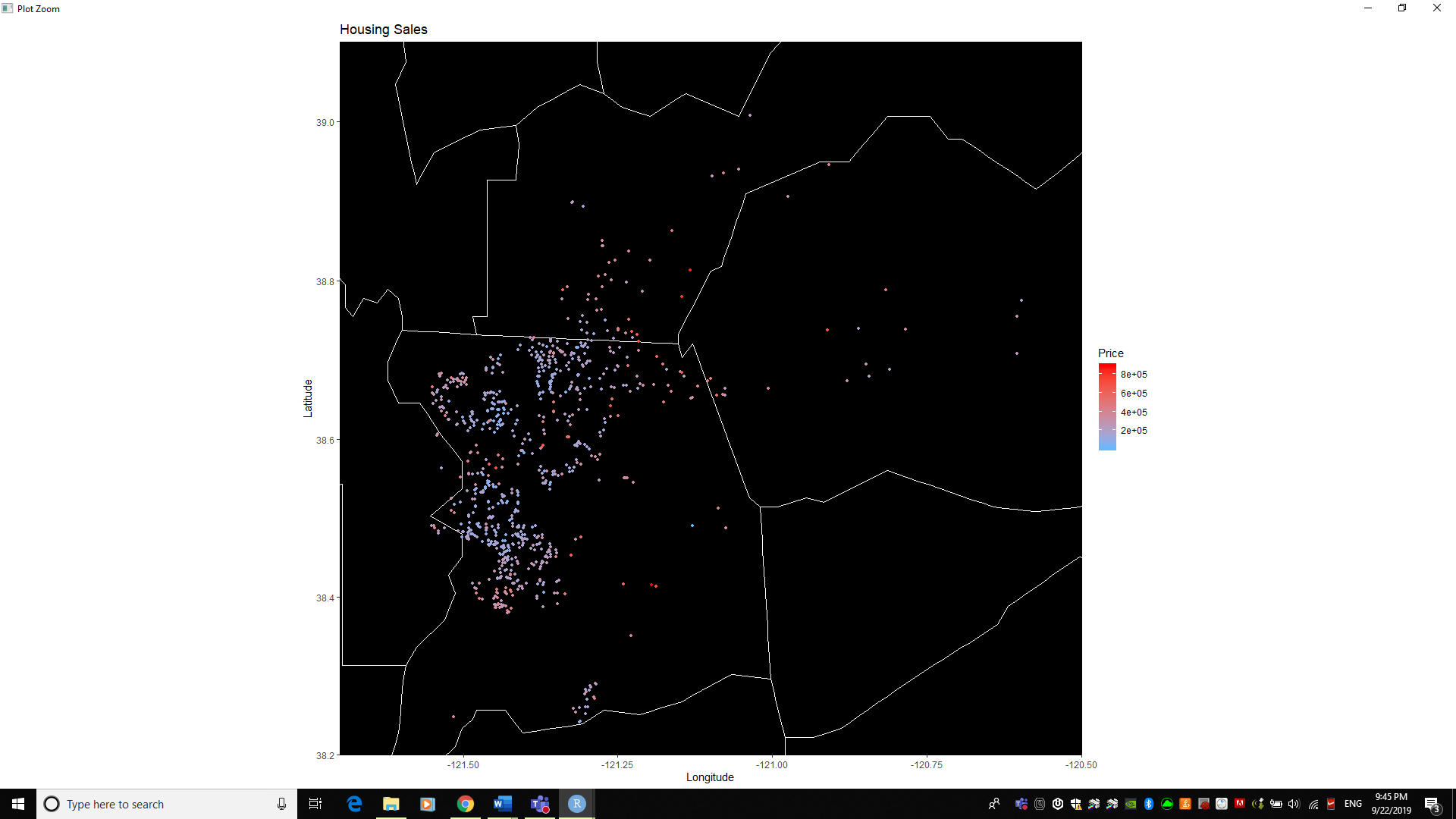


Fig 4) Geographic Representation of Housing Prices

We plotted each home’s location on the map with the shape’s fill set to cost. Our intent was to identify pockets of wealth and subsequently higher-priced houses that might factor into answering our questions. While we realized that the majority of our homes were located inside the city lines of Sacramento, we could clearly see pockets of higher-priced homes in certain areas. Our map overlay was basic, simply outlining county boundaries and did not provide us any more insight into the reason why these homes were priced above the rest. To better understand geographic and socio-economic circumstances that contributed to housing price, we would need to understand livability indices, crime rates, education quality and other factors. This was outside the scope of our project, but showed us how important it is to capture all relevant data, and reinforced the notion that a data scientist should consult subject matter experts with extensive domain knowledge prior to any project.

Once our EDA was complete, we utilized a series of models, including linear models, kernel support vector machines and support vector machines to predict housing price. For the first model, we used the variables of “# beds”, “# baths”, and “square feet” to predict price. When we observed the p values for our independent variables, we noticed that square feet, and number of bedrooms were significant, while baths was not significant. Our adjusted r-squared value was .477, meaning that ~48% of the variation in our dependent variable was explained by variation in our independent variables. This is a point where our team learned that “more predictors is not always better.” Reflecting on this project, with the inputs we had used for our model, we should have dropped “number of bathrooms” from this model because its P value was .21, which is not statistically significant. We found this model to be a decent predictor of home sale prices for single family homes, however, when applied to multi-family homes, the model itself was not statistically significant with an F-statistic p-value of .84. After training and testing with the kSVM and SVM models, we determined that the linear model had the smallest RMSE, and was the best model to use.

*Results*

In completing this project, we were able to answer our initial business questions but gained broader insights as well. We learned that the overall price-range for properties was from $2000 – $884,790, which illustrated how outliers can skew your data. With such a wide range for housing prices, we also had to consider that there might be other more “intangible” factors affecting housing price, such as location, view, crime rates, foreclosures, etc. It is very difficult to attain this granular level of detail when dealing with such a large dataset, which is why it is imperative for a data scientist to study the subject they are analyzing rather than make broad judgements based on patterns in the data; Contextual knowledge is key.

The client was interested in purchasing a 5-bedroom home and inquired about the price range in the area. Our response concluded that the average price for a five-bedroom home was $390,501, but that the standard deviation was quite high, suggesting that not all five-bedroom home were alike, which is intuitive. At this point we should have realized that we required more information from the client on the specifics of what they were looking for and their price range, given the wide range of possibilities. There is a stark difference between purchasing a 5-bedroom home with a scenic view, versus a foreclosed home that needs complete remodeling.

As a final question, our clients understood the price variability and asked if there was potential for them to purchase a 5-bedroom property while still having enough money left in their budget for a small condo for their college-aged children. We explained that they could find a well-suited home under their budget because of the price fluctuations but insisted that they look at homes in the $300,000 price range to assess the need to perform home improvement projects before they consider purchasing two properties. It would be unreasonable for us to suggest purchasing two homes before the client decided that the neighborhood and amenities in their first home was acceptable.

*Conclusion*

For an introductory class and project, this work was extremely valuable as a first-step into the data science field. It provided us the opportunity to make mistakes in analysis while learning the programming tools and methods necessary to build upon in future projects. We learned to leverage each team member’s strengths, whether it be presenting, programming, organization or team-leading and combined those skills with newly learned technical skills to approach a realistic problem set and provide thoughtful recommendations to a client. What was particularly valuable to me was the learning how to apply data visual tools to approach a problem from another angle. As a predominantly visual learner, I had always struggled with abstract or high-level concepts without a concrete visual to summarize the data. This provided a foundational ability to truly explore data and visualize it through different methods, to recognize patterns and build models based on information that we might miss without these skills.

**Case Study 2**

*Introduction*

Data management is a hotly-debated issue in the United States Military. The Department of Defense (DoD) has contracted multiple private organizations to develop systems in support of specific end-user business processes, ranging from intelligence collection to communication systems, program management and healthcare record management. A few of these organizations and systems include: Microsoft, Raytheon, Boeing, Lockheed Martin, BAE systems, Palantir, the Army Healthcare Longitudinal Technology Application (AHLTA) and the Combined Information Network Data Exchange (CIDNE). With so many organizations supporting the DoD, there are issues of interoperability and integration. This is true even in what should be the simplest field: database management and record keeping.

In my career, I have become familiar with multiple platforms for information management, including shared drives from Microsoft Windows, share portals, CIDNE and Palantir. Not only do these systems not communicate with one another, but each has their own administrative support network, requires different credentials and can be disrupted as users move between organizations. This is quite a problem for an organization that sees itself as a future leader in data generation and consumption

In 2016, as an operations officer for United States Indo-Pacific Command (US-INDOPACOM), I began to recognize that the DoD’s ability to manage programs, assess their effectiveness and plan for future implementation was significantly degraded by ineffective data management processes. It was not until I took IST 659 Database Administration Concepts and Management did, I realize that some of these issues could be improved if stakeholders came together to discuss the needs of users on the ground, and analysts – and develop a platform to support their operations. For the project in IST 659, I decided to generate imitation data and use it to assess the effectiveness of fictitious information operations in the European Theater. My intent was to model the interrelation of tactical influence operations on broader strategic objectives. The idea for this project came from my realization that we often conduct activities without a clear linkage to overarching objectives, which essentially leads to lost opportunities. This is particularly salient repercussions in the European theater, where state and non-state actors are spreading disinformation, with intent to fracture the North Atlantic Treaty Organization (NATO).

In my scenario, the Commander of the United States European Command is concerned about NATO efforts to counter adversary information operations because of restrictions limiting the number of allocated forces. The commander requested an Information Operations Command (IOC) to develop courses of action and to assess the effectiveness of those operations in achieving high-level objectives. The IOC has decided to maintain a database that establishes baseline data, tracks all operations and measures task completion and assesses effectiveness. At the end of the project, our goal was to determine:

* Which Countries were most receptive to United States Government (USG) information operations?
* What Socio-economic, diplomatic and developmental conditions are necessary for USG information operations to thrive?
* Which type of information-related capability is most effective and to what can we attribute that success?
* Are our priority information operations objectives sufficiently supporting the commander’s priority objectives?

*Methods*

To answer these questions, I designed a database with 18 tables that linked a target audience in a particular country to USG objectives through tactical implementing units.

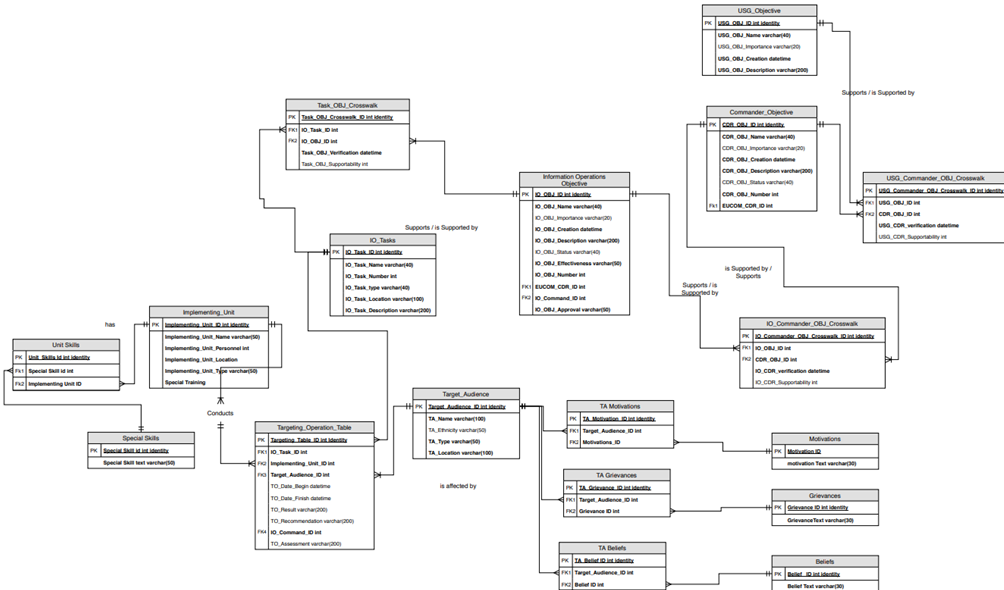


Fig 5) Logical ERD for Information Operations in Europe

From there, I considered each units’ operation report to be its own entry in the operation table. This table would link implementing unit, target audience, task, time of task, country and assessments to high-level objectives and ostensibly, allow us to answer the aforementioned business questions. After creating each table, I inserted data into these them via a combination of excel spreadsheet import and direct insertion. I generated multiple different categories of units, set their locations and attributes, special training and assigned them to one or more countries. For the target audiences, which are groups of people living in various European countries, I updated their beliefs, motivations and grievances by linking multiple many-to-many relationships between tables. After this was complete, I could get a holistic look at a specific group and in the future, assess the effectiveness of operations against groups with certain motivations or drives.

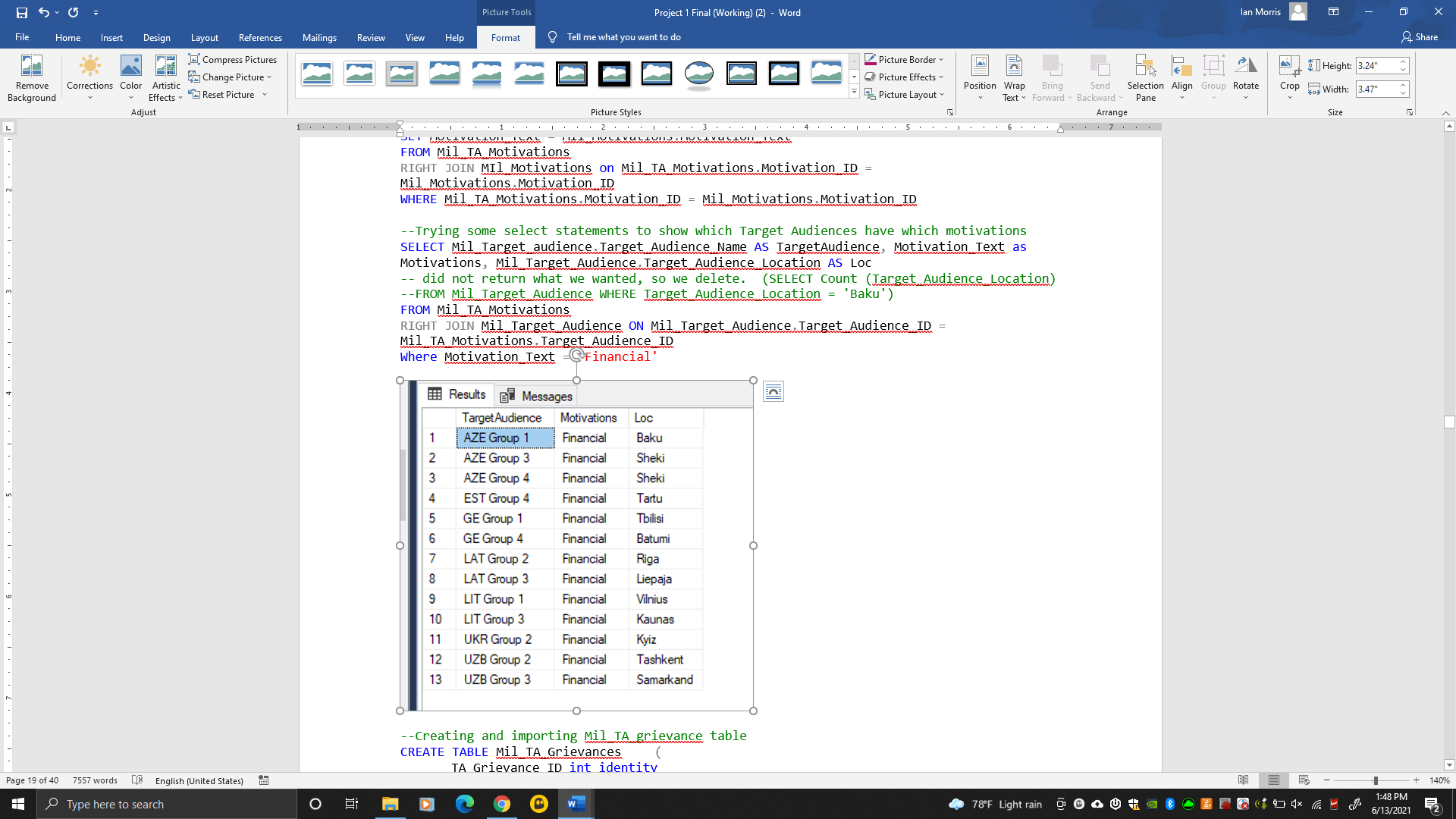


Fig 6) Target Audience Motivations

Finally, I linked the series of regional commander objectives to the U.S. Government objectives and identified how many tasks were directly supporting the top priority tasks and their overall support indices.

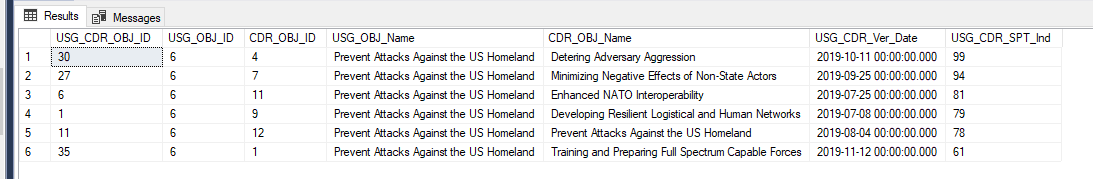


Fig 7) U.S. Government Objectives Support

*Results*

To answer my questions of which countries were most receptive to USG information operations, I grouped all of countries by how effective our operations were against populations within that country. There were over 350 distinct operations in my database and I averaged the efficacy of those operations to get a final score for the country. I accomplished this by joining the target audience table with the operations table. What I found was that certain countries had a higher “success rating” than others. I wanted to know if this was because of the environment, the population itself or the type of activity being conducted. In real-life implementation, it would of course be a combination of these factors, but for the sake of this project, I tried to identify a particular cause of success. I compared the most receptive country (Estonia) to the least receptive (Azerbaijan) in order to identify differences in operations between them. While I could not identify any significant differences in operations, I could examine at the attributes of the groups in both countries to see what it means to “be” Estonian and Azerbaijani.

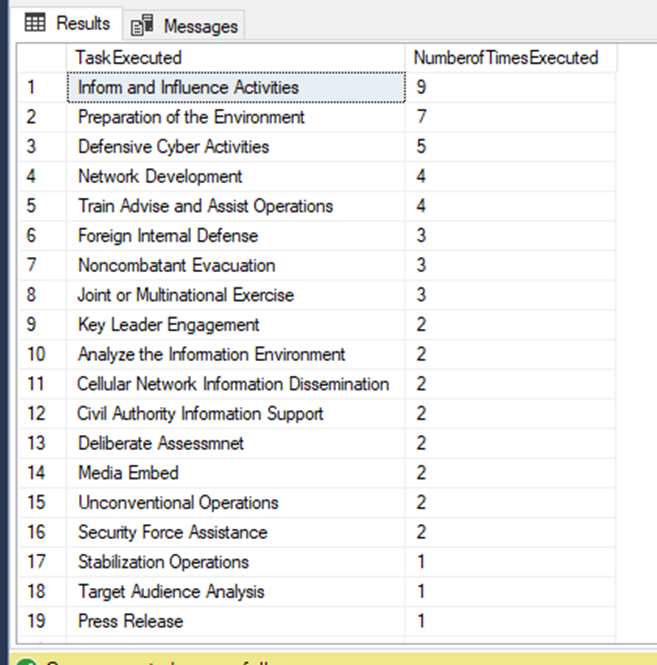
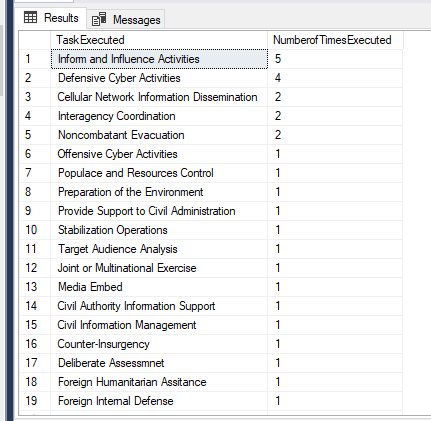


Fig 8) Executed Activities in Estonia (L) and Azerbaijan (R)

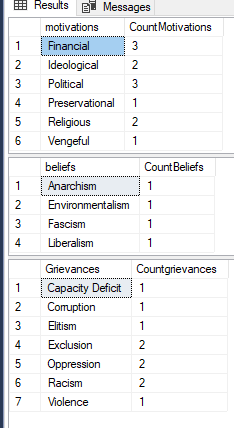
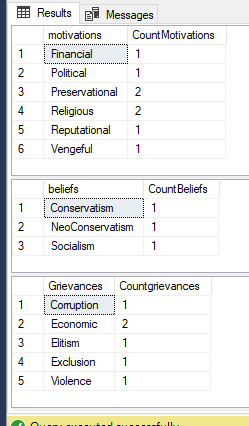


Fig 9) Group Identities in Estonia (L) and Azerbaijan (R)

When I viewed each country’s motivations, beliefs and grievances, I noticed that the least receptive country (Azerbaijan) had the same number of grievances as Estonia, but more individual groups and many more grievances. This led me to believe that Azerbaijani target audiences are more difficult to influence because they could not be unified beyond a driving cause. In this case, that hypothesis itself is not enough to make decisions, but rather, we should bring these insights to a multi-disciplinary team and allow them ascertain the nuance.

*Conclusion*

Truly effective organizations need to utilize database management systems to ensure the integrity, confidentiality and accessibility of their data. These attributes ensure rapid and informed decision making at the end-user level, maintaining initiative in operations. As data generation and consumptions increases exponentially, it becomes imperative that organizations set a healthy foundation from which to build. Prior to studying the concepts and completing this project, I had very little understanding of how we could manage data across teams, organizations and geographical locations. In completing this project, I discovered a new way to understand entity-relationship models, normalization and how to perform operations to assess effectiveness and answer business questions for end users.

**Case Study 3**

*Introduction*

My ultimate goal in undertaking this program of study was to learn new and innovative ways to define a problem set, explore the available data, and utilize models and statistical techniques to improve my organization’s approaches. My data analysis and decision-making course was the first opportunity to explore new tools through a data-centric approach and learn how to apply those to my organization through a structure framework. Often in studying data science we can lose our path when discovering different algorithms, programming languages, and ways to visualize data. This can be to the detriment of our ultimate goal of problem solving which is why I appreciated the structured methods presented of this course.

In this case study, I conduct a process improvement project to increase my exercise time and intensity following a shoulder injury and repair a few months prior. I chose this project topic after reading a study about the Army’s new fitness test that intends to reduce musculoskeletal injury and lost work time. My physical therapist at that period recommended that I maintain a 40+ minute daily workout regimen to expedite the healing process and was concerned that I wasn’t consistently reaching that goal. Overall, my objective was to increase workout time by 20% while maintaining 115 heart beats per minute during each exercise period. I would complete this over a two-month duration. I use the DMAIC approach: define, measure, analyze, improve, control to accomplish this project, which illustrates that you can apply these business processes to any type of problem-set in order to effect change. COVID-19 lockdowns compelled our class to take unique approaches to our process improvement topics which reinforced that the proper data collection and analytical methods can yield thought-provoking results.

*Method*

I utilized the aforementioned 5-step DMAIC model which began with the define phase. Here I mapped my pre-modification daily routine to determine which activities to consider that might be affecting my baseline exercise duration and intensity.

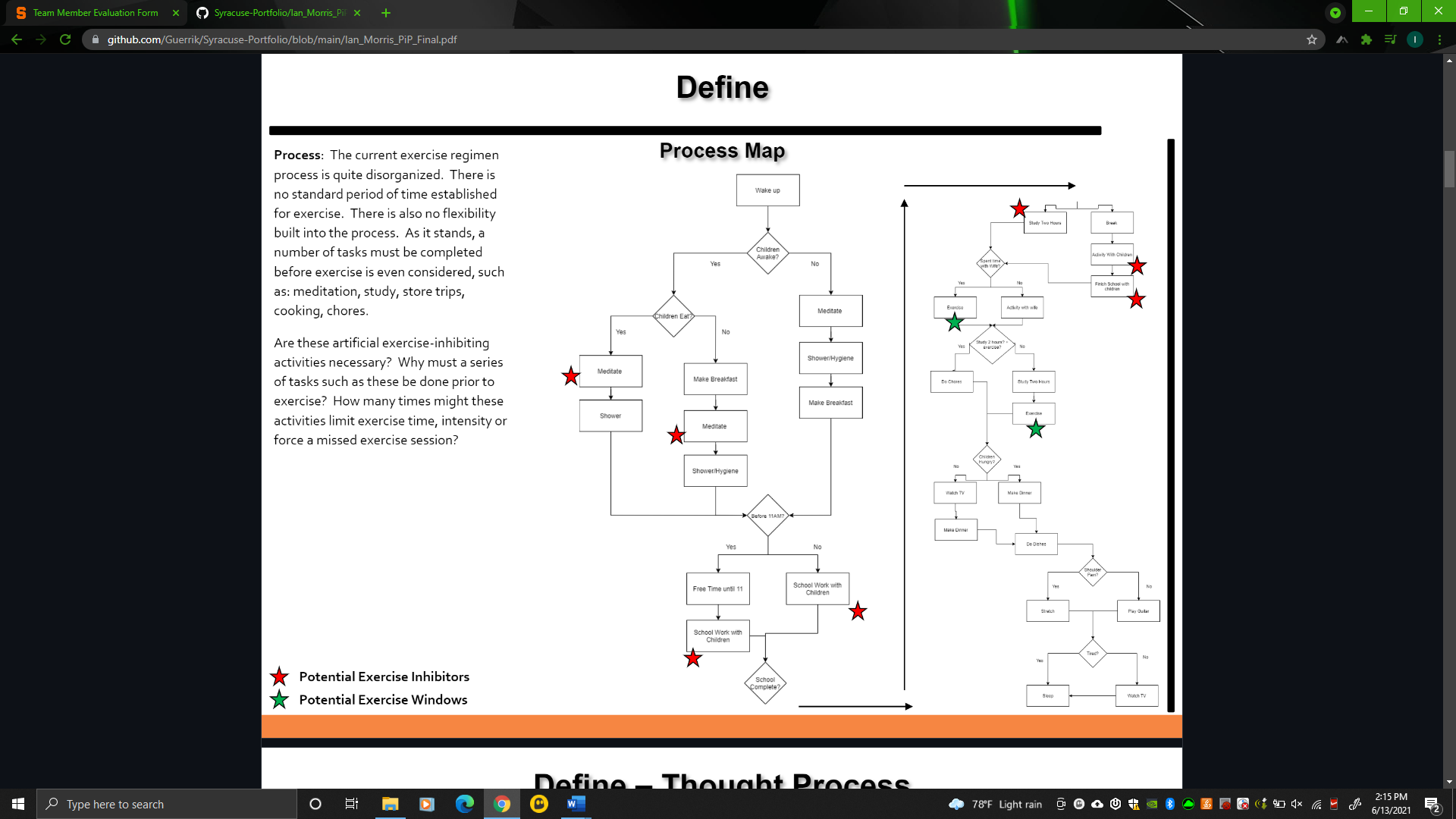


Fig 10) Process improvement Initial Map

I was able to identify inconsistencies in the time of day that I conducted exercise, likely inhibitors, and slow periods in each day where I could find additional time to exercise. This process map drove me to creating a “thought process map” where I would consider potential drivers of exercise time and any logical factors that I could uncover that might be shifting exercise time. I contemplated changes to these factors and the externalities that might arise from doing so. These included: water intake and diet, pacing exertion throughout my exercise, video games and other leisure, anxiety/meditation and work commitments. For example, I considered that anxiety might be resulting in less desire to exercise, and therefore I needed to record my daily meditation duration and the number of hours of schoolwork or work I had completed in that day. For hydration, I decided to record the amount of water I consumed versus caffeine. The belief is that if I can determine intuitive expected drivers of exercise time, I would be more likely to choose the correct information to track in my experiment.

In the Measure phase, I planned how I would monitor my process before any changes. I collected both continuous and discrete data which included things like: sleep time, water drank, stretching duration, study/class time, activities with my family, fruit consumed, shopping outings, weekend vs. weekday. Once I determined what I would be collecting, I needed to determine the collection method. I chose to record data daily, with a specific set of exercise activities to perform. I chose to use a heartrate monitor and multiple phone applications to measure my heart rate and exercise time from start to finish.

I concluded that based on my anticipated 5% margin of error at 95% confidence level, I would need a sample size greater than 385 days. I only had enough time to collect 32 observations during the “baseline’ month, so this meant that our smaller sample size would increase my overall margin of error. After collecting my 32 observations, I recorded a total of 25 defects (out of 64 possible) where I did not achieve my standard goal of 40+ minutes of exercise or average heart rate of 115 BPM. Translated into a Sigma Quality Level (SQL), this results in a score of 1.8 or yield of 61.8%, which left extensive room for overall improvement.

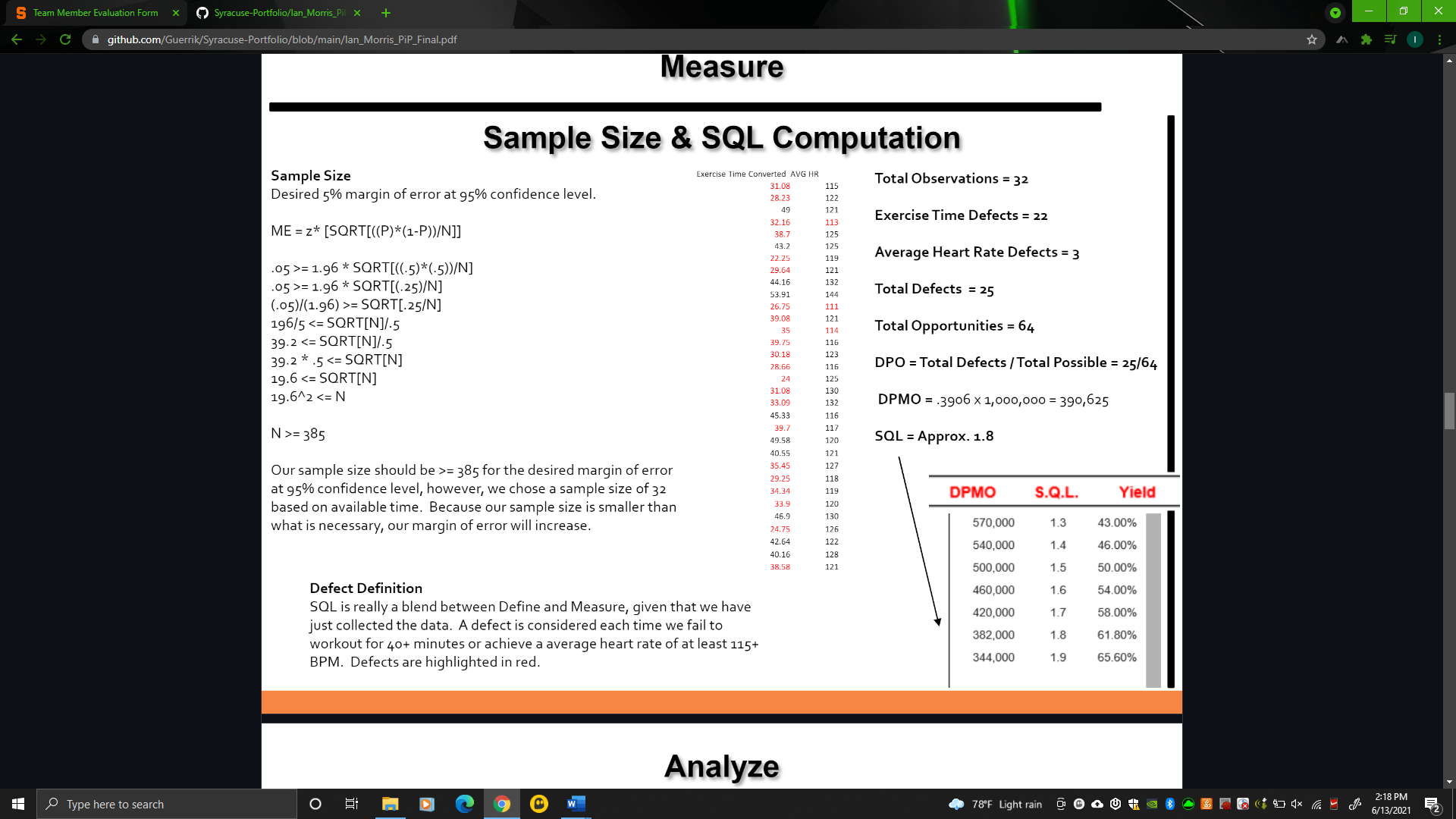


Fig 11) Sigma Quality Level Assessment

In my analyze phase, I utilized a series of exploratory data analysis techniques, including data summary graphs, correlation plots, chi-square tests, and multiple regression to determine if my chosen independent variables had the impact on my dependent variables that I expected. From the onset I noticed that my average exercise time fell at 36.28 minutes per day with a significantly large standard deviation of 8 minutes. This indicated that my daily exercise duration was extremely inconsistent. With the average duration less than the intended duration, I noticed that I was constantly underperforming in duration about 60% of the time.

In order to search for the root causes I created a correlation plot which measured the interrelation of the variables I expected to have an impact on exercise duration and heart rate.

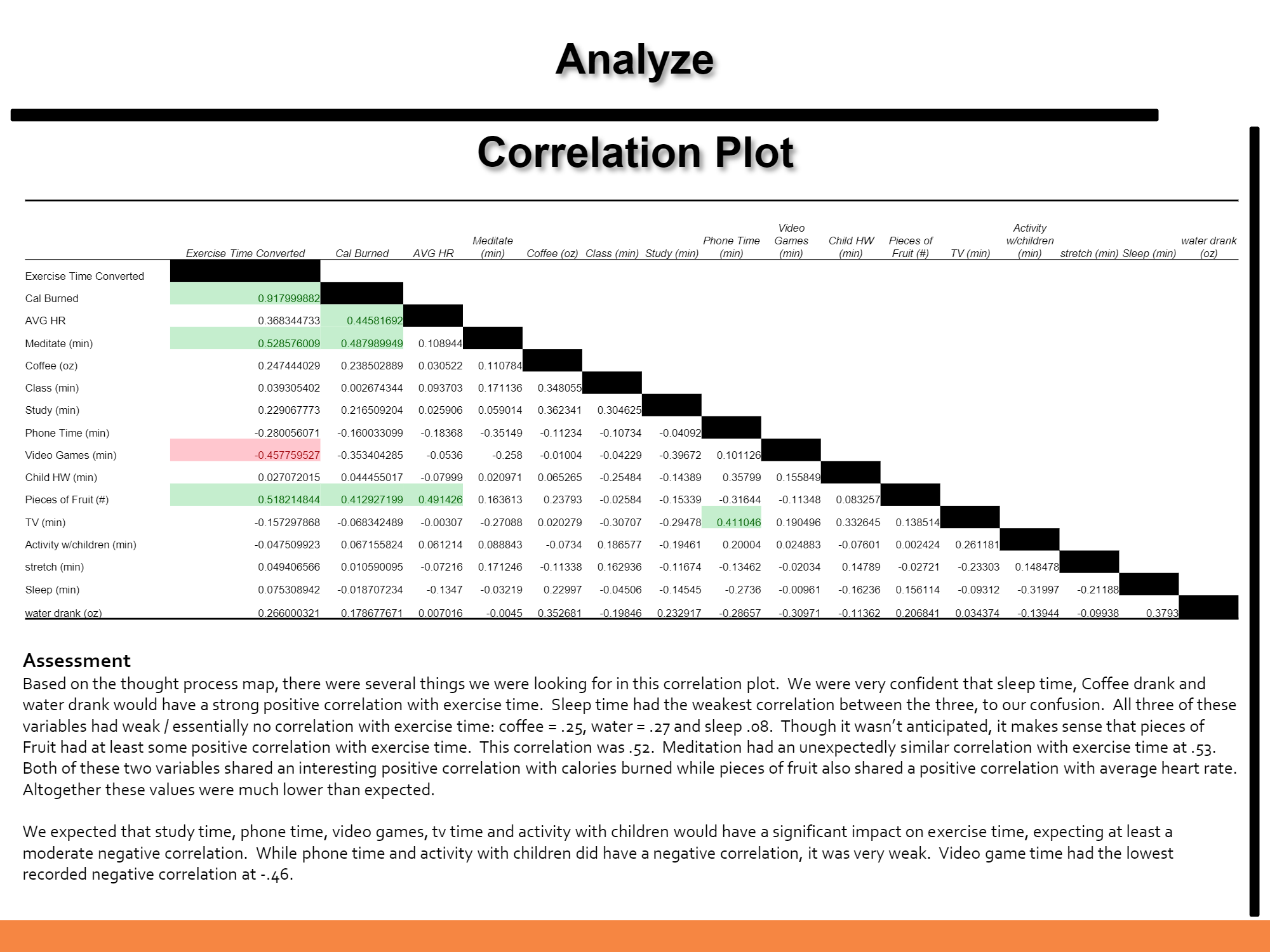


Fig 12) Correlation of Variables

At inception I believed that sleep time, coffee drank and water drank would have a strong positive correlation with exercise time as I believed that as they increased, my energy level would increase drastically. To my confusion, I observed that neither of these three variables had any significant correlation with exercise time. What I did observe in this plot was that meditation and pieces of fruit consumed had this largest positive correlation at .53 and .52 respectively while ‘video games played’ had the highest negative correlation at -.46. This illustrated that our instinctive choices may be weighted with bias and not have the impact on our dependent variables that we expect. After determining three potential drivers of exercise time, I ran a multiple regression to ascertain their effects. All three of these variables had p-values that were below .05 and significant and my adjusted r-square was approximately .51.

I would target these variables during the improve and control phases to test if they had a positive effect on my exercise time. I increased fruit intake from an average of 2.8 pieces per day to 5.5 pieces per day. I reduced video game consumption from an average of 73.6 minutes per day to 42.5 minutes per day and increased meditation from an average of 9.2 to 13.7 minutes per day. Using the regression equation from my multiple regression, I anticipated that a day where I ate 6 pieces of fruit, played 45 minutes of video games and meditated for 15 minutes would yield an exercise duration of 50.3 minutes.

I developed control charts to track my pre / post modification exercise times to monitor how much my process was varying and if it was occurring predictably.

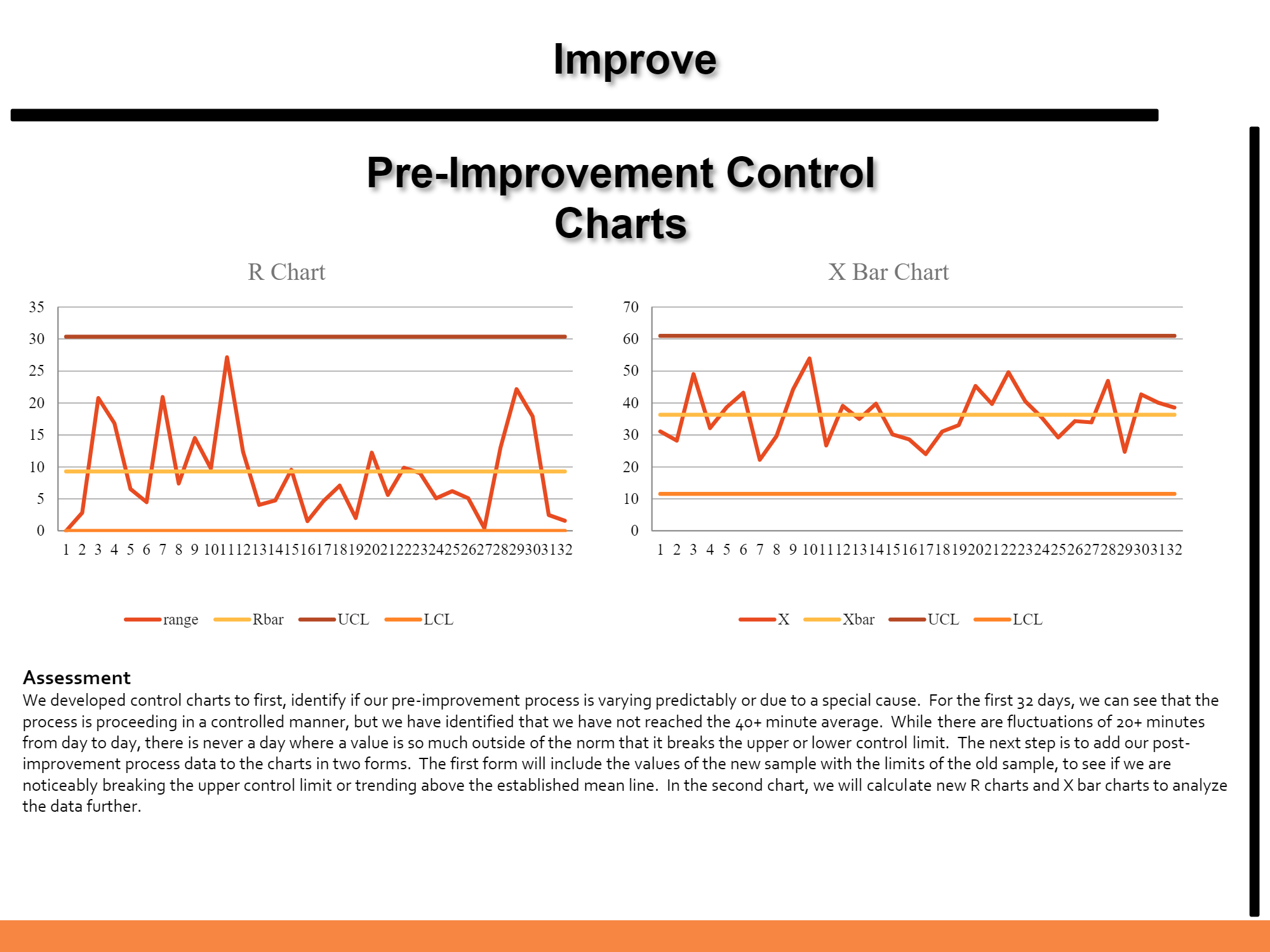


Fig 13) Pre-Improvement Control Charts

For my pre-modification data, my exercise times varied normally with no significant anomalies or patterns. My next step would be to append my post-modification data to examine whether the changes to fruit intakes/video games/meditation resulted in exercise times that broke my control chart limits and implied a noteworthy improvement. After two observations of normal variation, I witnessed exercise time variation that broke the upper control limit and maintained numbers above my average line for the next 10 points, denoting significant improvement in exercise time. I then calculated my post-modification SQL score which included 5 defects out of a total 32 opportunities. This resulted in a defects per million opportunities of 156k, or a SQL score of 2.5 with a yield of 84.20%, a substantial step up from our initial SQL of 1.8.

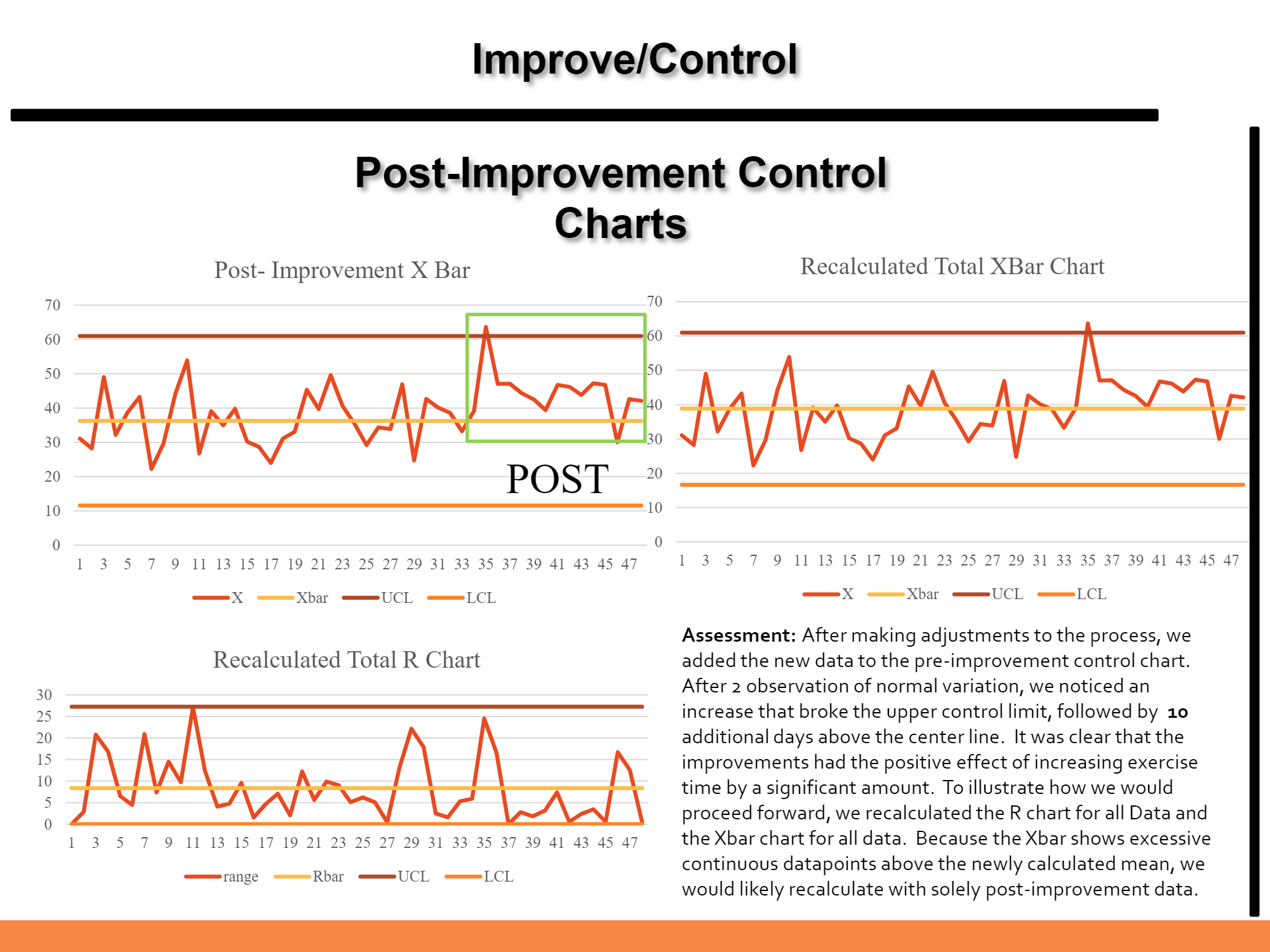


Fig 14) Post-Improvement Control Charts

*Conclusion*

What I learned from this process improvement project is that it can often be challenging to identify the root causes of a particular success or failure. This project was important to my development because it allowed me to work through an entire process from identification of the problem through implementation. I learned how to examine the problem in depth, generate a thought process and potential causes, test my hypotheses, isolate potential drivers and control future steps. I also learned that it is difficult to be a bystander when you have a vested interest in a particular conclusion. This is a great lesson for future team projects because it shows the importance of being an honest broker and refraining from influencing the data or collection methods which could result in significant bias. I understand that there may be 3 or more ways to improve a process: improve the system, distort the data or distort the process. While we always seek to maintain the integrity of the data, there is potential for us to subconsciously distort our processes when we have a particular goal in mind. In my project, I may have subconsciously manifested this in my workouts by slowing my exertion to net a few additional minutes to meet my goal or taking longer breaks in between exercises to achieve the same goal. Additionally, in my pre-modification period, I may have stopped a workout when I could have continued on so that I could achieve to get a larger % increase in my post-modification period. We must also take care to ensure we are collecting the right data and standardizing certain variables in our study – for this project food intake, and the specifics of the workout. These are factors we must always be aware of in a process improvement project, and adequate safeguards and checks should be in place to ensure that our procedures are not compromised.

**Conclusion**

These three projects are just a few in a series of that taught me creative and effective ways to collect, explore and model data to achieve a business objective. Ultimately my reason for joining the ISchool’s Applied Data Science Program was to improve my ability to link strategic objectives to ground-level project management and operations. I had observed too frequently the issues that arise from poor planning predicated on faulty or incorrect data collection. My intent was to understand how quantitative methods could serve as a guide that set conditions for experimentation and served as a ‘sanity’ check for some of our more qualitative hypotheses. These new methods will complement my previous knowledge and work experience and enhance my ability to understand the operational environment and visualize it for my seniors. Paramount to this is the necessity to seek domain knowledge from stakeholders and those familiar with the particular field we are analyzing then apply the tools we discovered in this program to render a judgment about a field, derive new insights and guide key decision makers.

In our Housing Market case study, we learned how to apply these methods to a potential future situation to suit a customer requirement. This involved cleaning data, evaluating the veracity of the data, combining datasets, applying key statistical principles to avoid pitfalls and faulty analysis, overlaying data graphically and altering our approach to data visualization to derive new insights. We also earned how exploratory data analysis can raise questions for future analysis, propel more granular data collection and challenge our assumptions about causation. Finally, we understood how statistical tools applied improperly can fail when we don’t account for outliers, missing data points and other anomalies.

Due to the extensive amount of data the U.S. Army generates and collects, I chose the Information Operations Database Management case study to illustrate how my knowledge gained in this program could be used to support real-world operations immediately. As it stands, the Army uses many divergent systems resulting in interoperability and integration issues at every turn. I generated this case study because I desired to model how an effective information operations campaign framework could look when developed with stakeholder buy-in at all echelons. I believe it’s imperative that ground-level users have candid dialogues with senior leaders to discuss information management. In this way, senior leaders will be able to clearly state their objectives and understand how this might relate to data collection and management. This dialogue will inform data architects and database managers in how they facilitate the storage, assessment, analysis and dissemination of this information, which I see as clear roadblocks in my field.

In my process improvement project, I was fortunate enough to revisit mathematical and statistical concepts that I had not practiced for many years. I appreciated the contextual nature of their presentation, namely that there was always purpose behind why we were learning the content and it wasn’t presented ‘in a vacuum.’ While the Army has its own decision-making process that has overlap with the DMAIC format, nearly all of our analysis is qualitative. Even when we discuss numbers, there is generally nobody in the room with the background knowledge to determine what the numbers truly mean, which can lead to faulty decision making with severe consequences. I look forward to my next planning conference where I will be able to guide my team through systematic analysis, improve and control phases as we address a wide range of problem sets, with specific focus on medical support to the Korean Peninsula. And 3-5 years down the road, I seek to dive deeper into the army’s Operational Research / Systems Analysis (ORSA) field which is a functional branch that leverages data science techniques, statistics, probability modeling and simulations to answer complex strategic business issues for some of the most senior leaders in the United States Department of Defense.