**Directions**

In your group you are expected to collaborate on addressing each question. If you are the first person to post your ideas, do not fear! This is intended to provide you with a safe space to put your ideas on paper.

If you come to the document and there are ideas that you disagree with or would like to revise / clarify, feel free to use comments to discuss with your group members.

My hope is that each group member contributes equally, but that may look different for each person. I expect for each of you to edit and add to posts from the rest of your group. This **is not** an individual assignment, so please don’t write individual responses and fail to engage with your group.

Your final responses are due on Friday, January 8 by 11:59pm. You will be required to download this Google Doc as a Word file to submit in Canvas. Only one submission per group is required.

**Question(s)**

1. What do you believe are similarities and differences between Statistics and Mathematics?

|  |  |
| --- | --- |
| Differences | Similarities |
| 1. Statistics tends to be more about using what we know to understand and make decisions about the future while the rest of mathematics is perfectly ok with remaining more abstract 2. Statistics likes using a lot of data 3. The answers of mathematics are usually certain (3+3 =6) while statistics can be uncertain (margin of error) 4. Statistics largely involves probability and determining whether a hypothesis is statistically significant for it to be assumed true. | 1. They both use numbers to answer and model questions we have about life 2. Both implement formulas to compute numbers 3. Statistics relies on Mathematics for computations |

1. From your perspective, what is Statistics? Is there a difference between “statistics” and “Statistics”? If so, what?

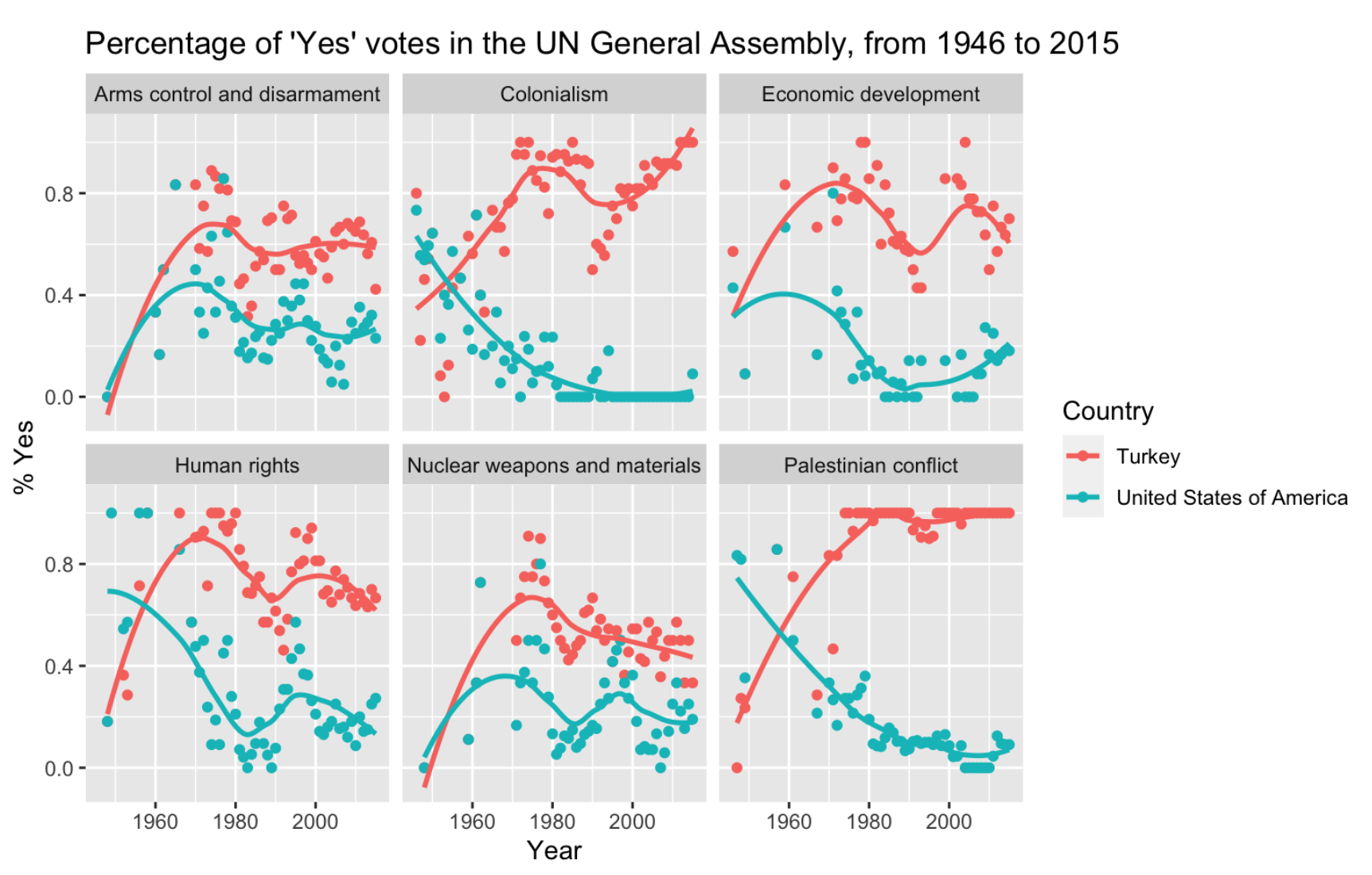
Statistics is about using the past to model and understand the future. We utilize data we gather from the past to make models to understand it so we can make predictions about the future. For instance, we can use statistics to determine whether a machine is faulty and needs to be replaced in the future. Statistics could be representing more of the general subject itself and data as a whole, while statistics might be more relating to the sample and its more specific characteristics, such as single measures of some value from the data.

1. Just because a variable has numeric values, does not mean it is a numeric variable. What are the different types of variables that can appear in a dataset? How can you determine if a variable is numerical versus categorical?   
   The different types of variables that can appear in a dataset include quantitative and categorical variables, as well as explanatory and response variables. These types of variables often have subtypes as well, such as discrete, continuous, ordinal, and nominal variables. You can determine if a variable is numerical if you can perform mathematical calculations on the variable, while you cannot do so with a categorical variable. These mathematical calculations that a numeric variable can undergo include addition, subtraction, and averaging. Even if a categorical variable is expressed in numerical terms, such as stages of cancer, the numbers do not have mathematical meaning. Taking the average of a measurement would confer a mathematical meaning to that value, whereas taking the average of the stages of cancer would not. For instance, a categorical variable could be the day of the week, while a numerical variable is something like the amount of residue left over after a reaction. Another example of a categorical variable is type of car, while a numerical variable could be height or weight because they represent numerical values.
2. What are the differences between observational studies and experiments?

In an observational study, the units have variable values that are not determined by the researchers. On the other hand, experiments are when researchers determine the values of one or more of the variables for observational units. Observational studies can be either prospective or retrospective in nature, while experiments can only be prospective. In observational studies, the investigators or researchers don’t assign the exposure in the group, they just observe what goes on and attempt to glean value from the observation without interfering. In experiments, researchers apply a treatment to a certain predetermined group, attempting to find potential causal connections between variables. Causation can be inferred in a randomized experiment, however, observational studies cannot reliably make this connection between variables. This is because in observational studies, there are confounding variables present that limit inference to association or form a hypothesis for further research, whereas experiments control for these confounding variables.

1. What are different methods for sampling from a population? Why would a researcher choose a stratified random sample instead of a simple random sample?

Two methods for sampling from a population are stratified random and simple random sampling. Stratified random sampling takes the original population and separates it into subpopulations based on similarities. Then, members of each subpopulation are taken at random for the sample. Simple random sampling does not involve any subpopulations. It is simply random sampling of the entire population as a whole. A researcher might want to choose a stratified random sample instead of a simple random sample because it allows the researcher to ensure certain aspects of the population are present in the random sample. A simple random sample could not provide the same guarantee. For instance, if a researcher is doing a study on association of exercise and grades, the researcher would want to use a stratified random sample in order to include students from different sports (baseball, soccer, tennis, football, golf etc) in the study. In some cases, a simple random sample could be optimal, as a stratified random sample is far more complex in terms of analyzing data. The researcher can also do cluster samples where they divide a population into subgroups and take a sample including all observations from a fixed number of those groups. Lastly, multistage sampling is another sampling method that is very similar to cluster sampling. However, instead of including *all* observations from the selected clusters, only a random subset of observations are taken from the selected clusters.

1. Describe what you see in the plot below. 

I see graphs on how many people from the US vs Turkey in the UN General assembly voted yes on certain topics over the course of multiple years.

Additionally, I see a positive correlation between the percent of Turkey’s population that voted ‘Yes’ in the UN General Assembly and the Palestinian conflict, as well as Colonialism. There appears to be a negative correlation between the percent of the US population who voted ‘Yes’ in the UN General Assembly and the Palestinian conflict and Colonialism. Additionally, there was no correlation between Turkey and the US’ percentage of ‘Yes’ votes in the UN General Assembly from 1946 - 2015 for economic development, nuclear weapons and materials, and arms control and disarmament. In general, there are less ‘Yes’ votes in the U.S. in comparison to Turkey. The amount of ‘Yes’ votes when it came to Arms control and disarmament, Nuclear weapons and materials, and Human rights, the graphs of the U.S. and Turkey are similar in terms of shape.