**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 |
| * In Statistics conclusions are not absolute while in mathematics there can be a definitive answer (2+2 = 4) * Math uses numbers to represent quantitative elements, In statistics numbers can represent counts of quantitative or categorical data * Statistics involves a greater degree of interpretation to a scenario, while pure math is not as applied. * Statistics incorporates many concepts that either contrast/are not seen in Mathematics, such as ethics, reason, context, randomness, inductive reasoning, etc. | * Both rely on mathematical concepts in order to find a solution * Solutions or calculations can be done using a calculator, computer or your brain. * Coding - like that in the R programming language - uses both statistical and mathematical reasoning to find its solution. * Both are practices that are necessary and have helped humankind make sense of the realities in our world |

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

“Statistics” is the study of statistical analysis as a whole, which heavily incorporates mathematics, analysis, data collection, ethics, interpretation, inductive reasoning, and many more concepts. Whereas “statistics” is simply a collection of data or single measure of data from a study that “Statistics” builds off of to use in different ways everyday. Furthermore, “statistics” are collected, then interpreted and analyzed through the study of “Statistics,” which lead to a conclusion(s), definite or indefinite. Then, those conclusions are presented and/or further studied through the collection of more “statistics.”

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 two variable types in a dataset are quantitative and categorical. The numerical variables include discrete or continuous variables. The categorical dataset includes regular categorical or ordinal. Categorical variables can be placed into categories, while numerical variables consist of a range of numbers to manipulate or take the mean of. Additionally, a simple way to determine if a variable is numerical versus categorical is determining whether engaging in simple arithmetic (averaging, subtracting, adding, etc.) is reasonable, if so, then that variable would be quantitative, or numerical. Moreover, it is crucial to identify the reasoning of how categories would be determined for categorical variables, when applicable. Furthermore, it is also important to indicate the measurement units, when appropriate, for quantitative variables.

1. What are the differences between observational studies and experiments?

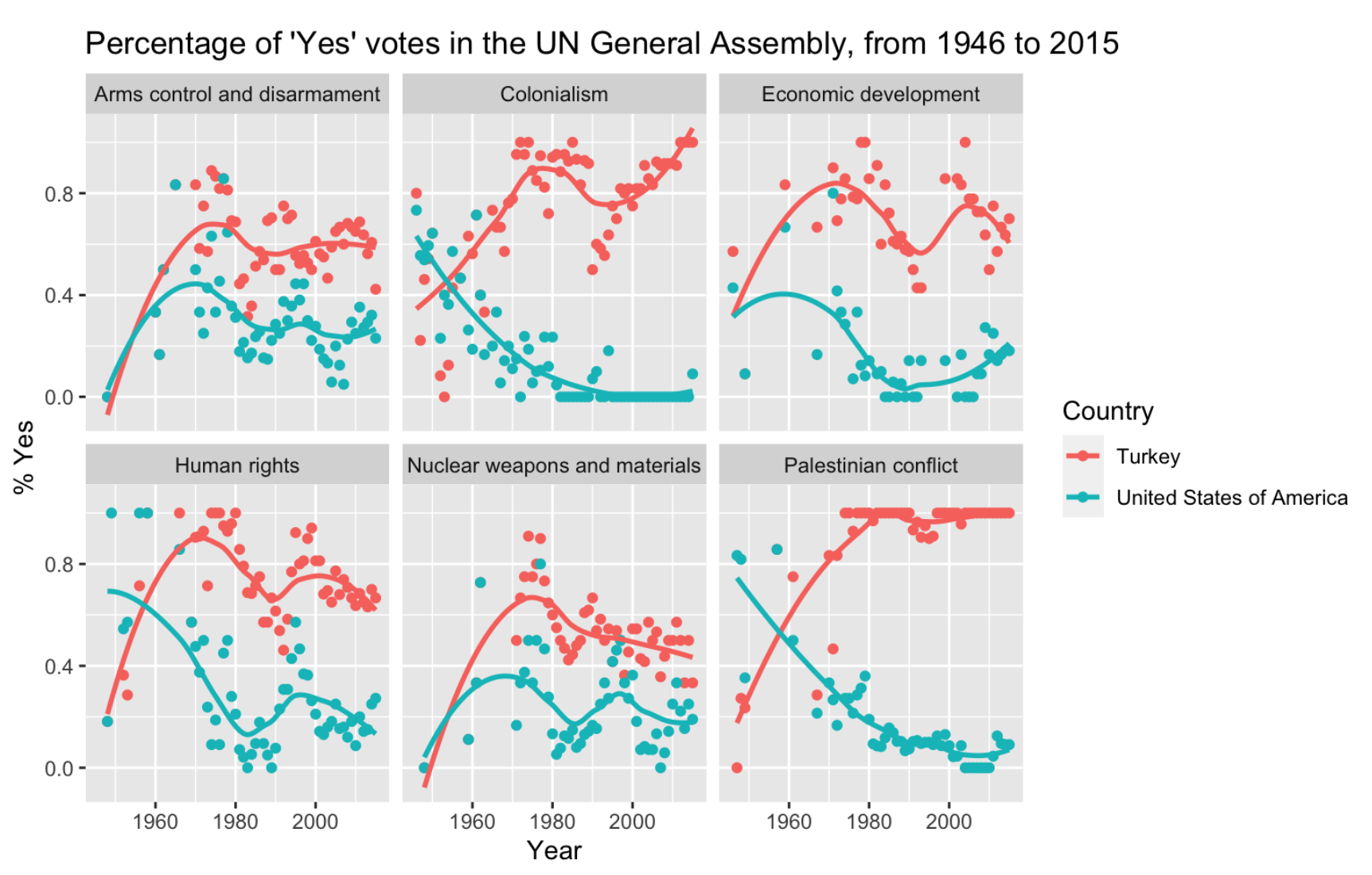
Observational studies are solely through observation and do not hinder the way that the data arises. In other words, observational studies is the collection of data that is recorded without the presence of control/manipulation. Furthermore, all observational studies have the potential to contain confounding variables.

Experiments are when researchers want to investigate causality between two (or more) variables and therefore impose one (or more) explanatory variable upon a response variable purposefully. Moreover, in experiments, groups are chosen, controlled, and manipulated by assigning the compared conditions. Common conditions that are identified and assigned in experiments are the control group and one (or more) treatment group(s).

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?

There are 4 different methods for sampling a population. The first is **simple random sampling**; as the name sounds, it is a simple random sample from a population. So in theory, a case would be selected at random, thus each case from the specified population has an equal opportunity/probability to be chosen and the sampling should be unbiased. The second is **stratified sampling**, or breaking up a population into meaningful groups or “strata” based on specific characteristics, then sampling from each strata. The third method is **cluster sampling**, breaking up the population into smaller groups called clusters and then selecting a fixed number of clusters and sampling all individuals within each cluster. The fourth method is **multistage sampling**, which is similar to cluster sampling however a random sample is taken from a selected cluster rather than collecting the samples of all individuals.

A researcher may choose a stratified random over a simple random sample if there are certain traits of groups within a population that are not equally dispersed in the population so that the researcher can divide the population into strata then randomly sample individuals within each subgroup. Furthermore, a stratified random sample allows groups within the population to have adequate representation. Stratified random sampling also allows for reduction in sampling variability and more precision in the information. Moreover, stratified random sampling is more advantageous when cases within each strata are respectfully similar to the interested variable. Therefore, choosing a stratified random sample would deem very beneficial.

1. Describe what you see in the plot below. 

The following plot was acquired through an observational study. Each country (Turkey and the USA) are a categorical variable. Furthermore, each individual year (i.e. 2000) is a categorical variable, while the “% Yes” is a numerical variable, specifically a continuous variable. Both, the individual year and each country are not ordinal variables. Moreover, it can also be inferred that the explanatory variable is the year of each assembly and the response variable is the “% Yes” votes in the assembly. In other words, the year at which the assembly is held may have impacted the “% Yes” votes from the assembly. From the plot, a general observation identified is Turkey has higher percentages of voting “yes” in the UN General Assembly from 1946 to 2015 on the issues listed, than the USA. Additionally, as time progresses towards 2015, the USA and Turkey have a higher magnitude of difference in the agreement on colonialism and Palestinian conflict. Based on the study, Turkey is predicted to be more likely to vote “yes” in the future in all the issues listed, than the USA, for at least a few years after the study (i.e. 2016-2020).