Kinlie Gohl

Brush up on R and Quarto

What problems do you encounter when working with the dataset?

> When working with the data set, I noticed that there were some missing data points. For instance, in highincome, votekmt_nm, and income. I also noticed that there are highly skewed variables like age. There is an overrepresented number of older adults compared to middle aged adults. There also seems to be duplicated results. There are many entries from different districts while some districts only have a couple entries. That may be because of the number of people in each district, or because of duplicated entries.

How to deal with missing values?

> To deal with missing values, I used a code source to clean up the data and change every missing value which was represented by -9 or 99, to N/A. I did so by running a summary to see which variables were missing using colSums(is.na(teds_2016)). Then the dplyr package changed all the variables.

Explore the relationship between Tondu and other variables including female, DPP, age, income, edu, Taiwanese, and Econ_worse. What methods would you use?

> To compare different groups like Tondu, female, age and more we would first need to find out what each variable is. For example, female is binary because there is only the option of 1 or 2, male or female; this is the same for DPP, yes or no. Taiwanese and econ_worse are also binary. Other variables like age and income are continuous are number based. Edu is an ordinal variable because it can be ranked by education level. Next, the method used would be based on Tondu's type. If Tondu was binary, I would use female and Tondu since they are both binary variables and run a bivariate analysis because there are two variables with only two options. A chi squared test would probably be the most useful for this variable interaction. To measure continuous and binary variables, like Tondu and age, I would use a boxplot because you would be able to see how the age differs when Tondu is yes and when it is no. This is also a good way to show outliers because points would show up way outside of the boxed area. Since edu is an ordinal variable, a chi squared test would show if there's an association or if the two variables are independent.

How about the votetsai variable?

> For votetsai, first we must identify what type of variable it is. It is a binary variable because it only has 1 or 0, did or did not vote for Tsai. A chi-squared test would be used to see if each variable is independent of each other.

Kinlie Gohl

Generate frequency table and bar chart for the Tondu variable.

	Tondu	Count	Percentage
	<fct>	<int>	<dbl>
1	Unification now	27	1.6
2	Status quo, unif. in future	180	10.6
3	Status quo, decide later	546	32.3
4	Status quo forever	328	19.4
5	Status quo, indep. in future	380	22.5
6	Independence now	108	6.39
7	NA	121	7.16