Assignment: Displaying univariate data

|  |  |
| --- | --- |
|  | **Unit 9 -** In this assignment you will learn how to describe one variable using a frequency table, a bar chart and a pie chart using SPSS. |
|  | [introduction SPSS](https://vimeo.com/user29453510/review/139201671/8ac05b25e8) (repeat)[Frequency tables](https://vimeo.com/user29453510/review/139203937/a0eac91269) (repeat)[Recoding and creating a variable](https://vimeo.com/user29453510/review/135956666/ef5cd1b5a4) (repeat)Youtube microlectures on building graphs in SPSS |  | Univariate analysisfrequency tablebar chartpie chart(geographical) mapHistogram\* time series |
|  | *Babbie, Ch. 14, p. 416-425* |  | [grouping] |  | 90 minutes |
|  | Read this assignment carefully and answer the questions. Bring either a print or digital version to the lecture/tutorial. |

The European Social Survey (ESS) is a large European data collection project that has collected data from all European countries every two years starting in 2002. The Prime Minister of the Netherlands Mark Rutte wants you to present some basic information about his citizens using the Dutch part of the ESS data. As he is a busy man, he wants all of the information to be very easy to understand, so it does not cost him too much of his time. Download the data matrix UNIT 9 - ESS7NL.sav and open it. Take a look at the variables in the dataset and consider their measurement levels.

1. Based on the measurement levels, what types of graphs or frequency tables are appropriate for univariate presentation? Do you understand why?

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Variable label** | **Measurement level** | **Type of graph/table** |
| tvpol | TV watching, news/politics/current affairs on average weekday | Ordinal | Bar chart (maybe pie) / frequency table |
| ppltrst | Most people can be trusted or you can't be too careful | Ordinal | Bar chart (maybe pie) (histogram if you consider it to be a scale) |
| polintr | How interested in politics | Ordinal | Bar chart (maybe pie) / frequency table |
| vote | Voted last national election | Nominal | Bar chart/pie/ frequency table |
| lrscale | Placement on left right scale | Ordinal | Bar chart (histogram if you consider it to be a scale) |
| rlgdgr | How religious are you | Ordinal | Bar chart (histogram if you consider it to be a scale) |
| gndr | Gender | Nominal | Bar chart/pie/ frequency table |
| agea | Age of respondent, calculated | Scale | Histogram |
| edlvenl | Highest level of education, Netherlands | Ordinal | Bar chart/ frequency table |
| emplrel | Employment relation | Nominal | Bar chart/Pie chart/frequency table |
| emplno | Number of employees respondent has/had | Scale | Histogram |

2. Can you use a histogram for a dichotomous variable? Why (not)?

Histograms are used for continues variables only.

3. Create a frequency table for the variable *polintr* (Political interest). There are three different percentages presented. Which one do you think is the most informative, if you want to present one percentage?

**Valid percentages,** it calculates the percentages for all the valid scores and the sum is 100%. Percentage including missing values is less informative, because you want to present the respondents who answered the question. It is important to realize that when the percentage of missings is very large it can have a great effect on the percentage you present. It also can have a great effect on the reliability and/or validity of the variable, depending on the reason. The cumulative percentage is least informative.

4. Create a frequency table for the variable *emplno* (number of employees the respondent has). This variable is only asked to respondents who are self-employed (i.e. had a score 2 on the variable *emplrel* (employment relation). Compare the column *percent* with *valid percent*. Think of a descriptive research question you cannot answer with the column *percent.*

For example: What proportion of people who are self-employed do not have other people working for them?

5. Create a bar chart for the variable *edlvenl* (highest educational level). Do you think this is very informative graph, i.e. is it quick and easy to understand? If not, what could you do to make it more informative?



Figure 1. Bar chart for highest educational attainment in percentages in the Netherlands in 2014.

A lot of values/attributes with long labels. Not really informative. Better to recode this variable into (for example) three attributes; lower educational level, middle educational level, higher educational level.

6. Recode the variable *edlvenl* into a new variable (educ\_cat). Recode education so that in the new variable educ\_cat has three categories (lower educational level, middle educational level and higher educational level). Don’t forget to give the variable a label and give the attributes value labels as well. Make a bar chart of the new variable. Has the bar chart improved?

Yes, It is easier to understand. The attributes are simpler, so you can have a quick glance at the chart and immediately understand the information.



Figure 2. Bar chart for lower, middle and higher educational attainment in percentages in the Netherlands in 2014

7. Sometimes, a pie chart is a good way to visualize a variable. Create a pie chart for the variable *vote* (did you vote in the last election) in percentages. Think about what you can edit in the lay-out of the pie chart to enhance the comprehensibility. Use the Editor to implement your changes. Think about what you want to show (did you vote), and what attributes you might not want to show. Recode if necessary.

The colors can be changed for example. They are now random, but can be changed into something more meaningful, such as green for the attribute *Yes* and red for *No* and grey for *not eligible to vote.* Also present a title, and add the percentages. You can also recode the attribute *not eligible to vote* as a missing variable, as you are probably interested in the voters vs. nonvoters alone.



Figure 3. Pie chart for percentage of voters in the last election in the Netherlands in 2014

8. For the last part of this assignment, we are going to convince the PM that there are a lot more Dutch women than there are men, as he is searching for a girlfriend and we want to give him a feeling that the odds are in his favor. Create a bar chart for the variable *gndr* (Gender). *By changing the scale of the Y-axis, create the illusion that there are in fact a lot more women than men.*

You can make differences blow up, by changing the range of the scale of the y-axis. By default it ranged from 0 – 60%, but if I change that to 42-58%, the difference between men and women is suddenly a lot bigger. If I than also get rid of the tick tables, you are left with a very misleading graph.



Figure 4. Misleading, but not per se fraudulous bar chart of the distribution of males and females in the Netherlands in 2014.

<<End of the assignment.>>