Intro to Variables and Data Transformation

Goals,

> Learn different types of data

> Learn basics of Power BI

> Perform required data manipulation

What is a variable? <https://www.youtube.com/watch?v=tHYis-DP0oU&t=19s>

Pay per hr > 10$ + [tips] = [tips] is a variable (can change from scenario to scenario)

Types of Data: Nominal, Ordinal, Interval/Ratio <https://www.youtube.com/watch?v=hZxnzfnt5v8&t=4s>

Data we collect are all forms of observations, variables record the measurements were interested in.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Variables | | | | |
| Observations |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Lvl of measurements, determines which summary statistics, graphs and analysis are possible and sensible

- Nominal (categorical and qualitative), can be given as a **frequency or proportions (percentage)** is used (mean or average value cannot be used to calculate this data)

Variables examples:

> Sex

> Preferred type of chocolate

> Color

**Graphed:**

>Pie chart

>*Column of bar chart*

>Stat column or bar chart

-Ordinal (have a meaning of order but the intervals between the values may not be equel), can be given as a **frequency and proportions (percentage), it is possible to calcucate the mean or average** for this type of data, however you have to know what your data means and is it justifiable

Variable examples:

> Rank

> Satisfaction

> Fanciness

**Graphed:**

*>Column or bar chart*

-Interval/Ratio (scale, quantitative, parametric), the most precise level of measurement, includes data that can be measured. Can be discrete (5 customers/ 17 points/ 12 shops) or continuous (4.2miles/25.3 degrees/3.8min). **Mean, median and standard deviation**

Variable examples:

> Age

> Weight

> Number of customers

> Size

**Graphed:**

>Bar chart

>Histogram (where data is grouped)

>Boxplot

>Line chart

(DataCamp, Data preparation in Power BI) What is clean data?

-Free from missing values

-Typos/data entry errors are corrected

-No duplicated data

-Irrelevant data is excluded

-Outliers are dealt with (cap them at a certain value)

-Correct data types for all columns

-Column & table names are short and descriptive

Cleaning data in power BI using Power Query > Data manipulation framework

-To load & transform data, we use the Power Query interface in Power BI

-Transformation steps are added sequentially

-Steps are saved in a recipe called applied steps that axecutes every time your data is loaded/refreshed

-To apple the transformations, we choose “Close & Apply” In the power query editor

Data types in Power Query

>Numerical Data (123, 14.45, 34%)

>Date/Time Data (01/01/2021), 9:48 AM

>Text Data (ABC)

>Logical Data (TRUE)

>Binary Data (REQDSdeqD=)