Introductory Statistics for Economics

ECON1013: LAB 1

Duong Trinh

University of Glasgow

Jan 2024

Intro

- Duong Trinh
  - PhD Student in Economics (Bayesian Microeconometrics)
  - Email: Duong.Trinh@glasgow.ac.uk

- ECON1013-LB04
  - Monday 1-2 pm
  - 3 sessions (29-Jan, 12-Feb, 26-Feb)

Record Attendance

Scan the QR code below or use the password listed below to take your attendance

9euzo3

Setup

- Step 1: Download Lab materials from **Moodle** page $\rightarrow$ Extract the folder in PC.
- Step 2: Log in **Microsoft onedrive** using your student account https://onedrive.live.com/login/ and upload the folder above.
- Step 3: Launch the **Excel** online https://www.office.com/launch/excel?auth=2, which we will use for all lab sessions.

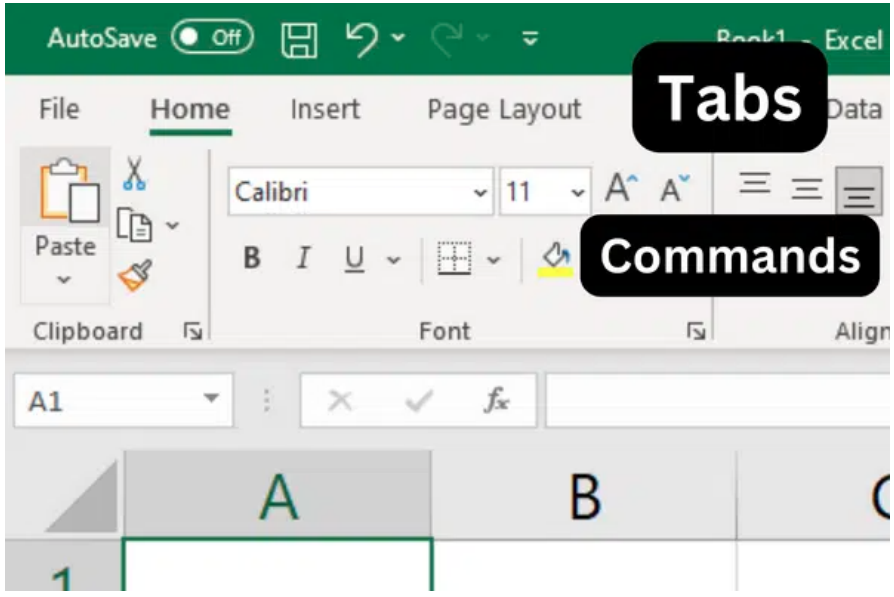
PRELIMINARIES

The Excel Interface

- Ribbon
- Workbook and Worksheets
- Rows, Columns, and Cells

The Excel Interface

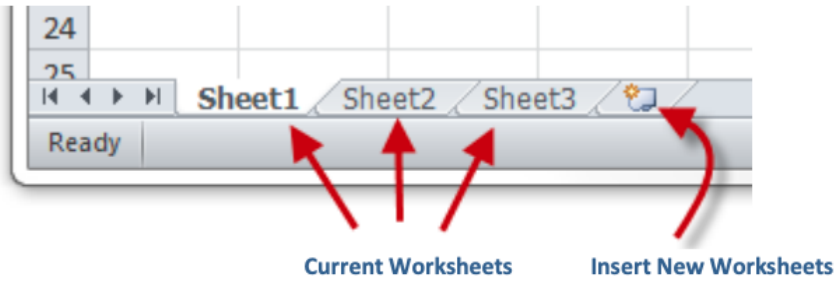
Ribbon



- To show or hide the ribbon commands, press **Ctrl-F1**.
- If you can't remember the location of a command, you can always use the search bar on the ribbon to find it.

The Excel Interface

Workbook and Worksheets

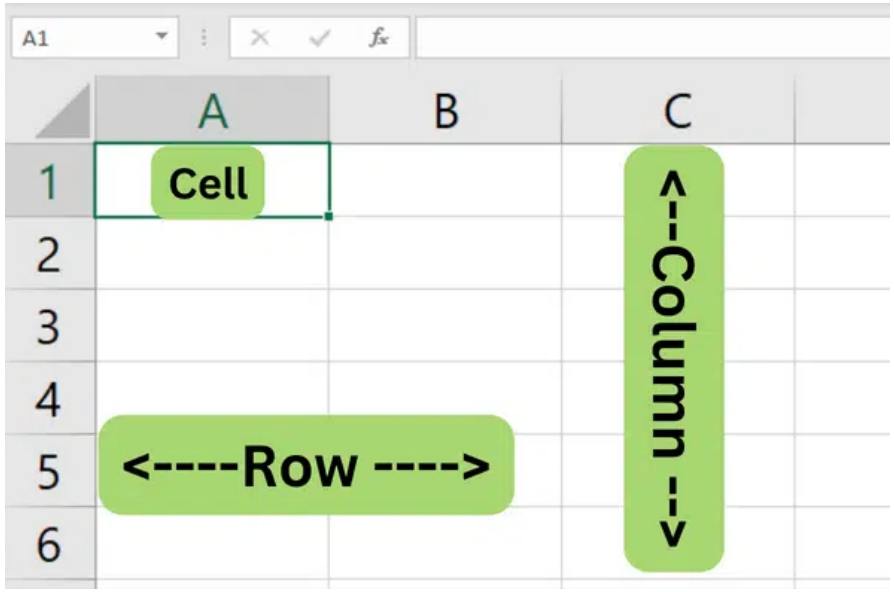


A workbook is an Excel file that contains one or more worksheets. Worksheets are where you organize and process your data.

- To navigate through your worksheets using keyboard shortcuts:
  - Press **Ctrl + Page Up** to move to the next sheet.
  - Press **Ctrl + Page Down** to move to the previous sheet.

The Excel Interface

Rows, Columns, and Cells



- To navigate within a worksheet, use the arrow keys to move up, down, left, or right.
- To select a range of cells, click and drag with your mouse or hold the Shift key while using the arrow keys.
- To quickly select the entire row or column, click on the row number or column letter.

Excel Formulas and Functions

Excel Formulas

Cell Referencing

Excel Math Functions

Excel Formulas and Functions

Excel Formulas

Formulas in Excel are used to perform calculations and manipulate data with built-in functions.

- To create a formula, start with an equal sign (=) followed by a combination of numbers, cell references, and mathematical operators.
- Here is an example that adds the values in the range `A1` to `A5`: `=SUM(A1:A5)`.

Excel Formulas and Functions

Cell Referencing

Cell referencing is a way to point to a specific cell or range of cells in a formula. There are two types of cell references: absolute and relative.

- **Absolute**: refers to a specific cell or range and keeps the same reference even when the formula is copied. It uses a dollar sign ($) to denote absolute referencing, like `$A$1`.
- **Relative**: A relative reference changes when the formula is copied to another cell or range, adjusting the reference based on the new location.

Excel Formulas and Functions

Excel Math Functions

Here are some commonly used Math functions for computation in Excel:

Function	Description
SUM()	Adds up a range of numbers
AVERAGE()	Calculates the arithmetic mean of a range of numbers
MIN()	Returns the smallest value in a dataset
MAX()	Returns the largest value in a dataset
COUNT()	Counts the number of cells containing numbers within a range
PRODUCT()	Multiplies a range of numbers together

Excel General Shortcuts

Here are some commonly used shortcuts for routine tasks and Excel commands:

Shortcut	Task
Ctrl + N	Create a new workbook
Ctrl + O	Open an existing workbook
Ctrl + S	Save the current workbook
Ctrl + Z	Undo the last action
Ctrl + Y	Redo the last action
Ctrl + C	Copy the selected cells
Ctrl + X	Cut the selected cells
Ctrl + V	Paste the copied or cut cells

- Shortcut to select data range: **Ctrl** + **Shift** + $\Downarrow \Uparrow \Leftarrow \Rightarrow$

Excel Charts And Graphs

Excel offers a handy variety of charts and graphs to choose from, including:

- **Column charts**: Compare different data sets across distinct categories.
- **Bar charts**: Display comparisons among discrete categories horizontally.
- **Pie charts**: Illustrate proportional data and percentages.
- **Line charts**: Show trends and patterns over time (aka *time-series plots*).

Excel Charts And Graphs

To create a chart in Excel:

1. Select your data range.
2. Click on the ‘`Insert`’ tab in the Excel toolbar and choose desired chart type.
3. Customize your chart’s design, layout, and formatting to meet your requirements.

Excel Charts And Graphs

To create a chart in Excel:

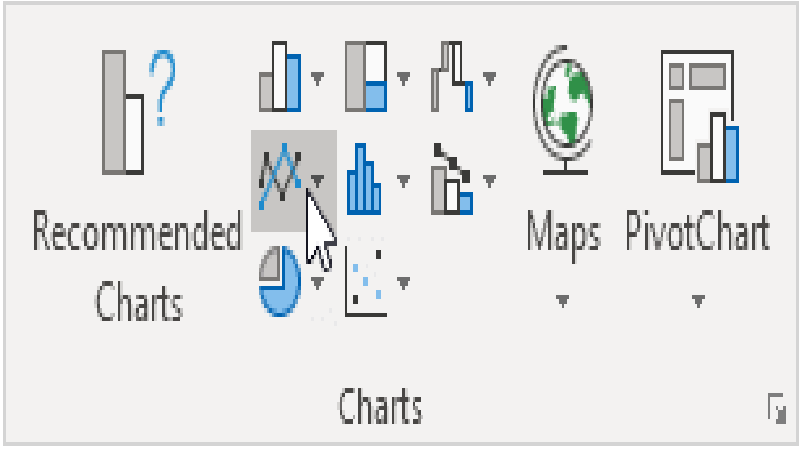
1. Select your data range.

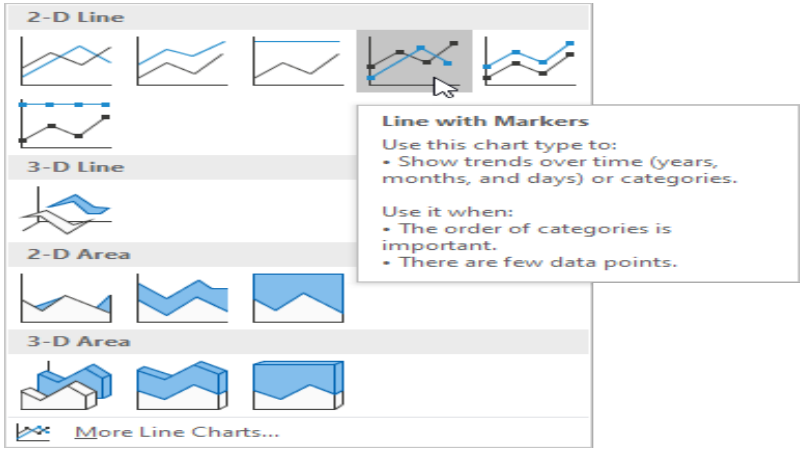
	A	B	C	D	E
1	Month	Bears	Dolphins	Whales	
2	Jan	8	150	80	
3	Feb	54	77	54	
4	Mar	93	32	100	
5	Apr	116	11	76	
6	May	137	6	93	
7	Jun	184	1	72	
8					

Excel Charts And Graphs

To create a chart in Excel:

2. Click on the '`Insert`' tab in the Excel toolbar and choose desired chart type.

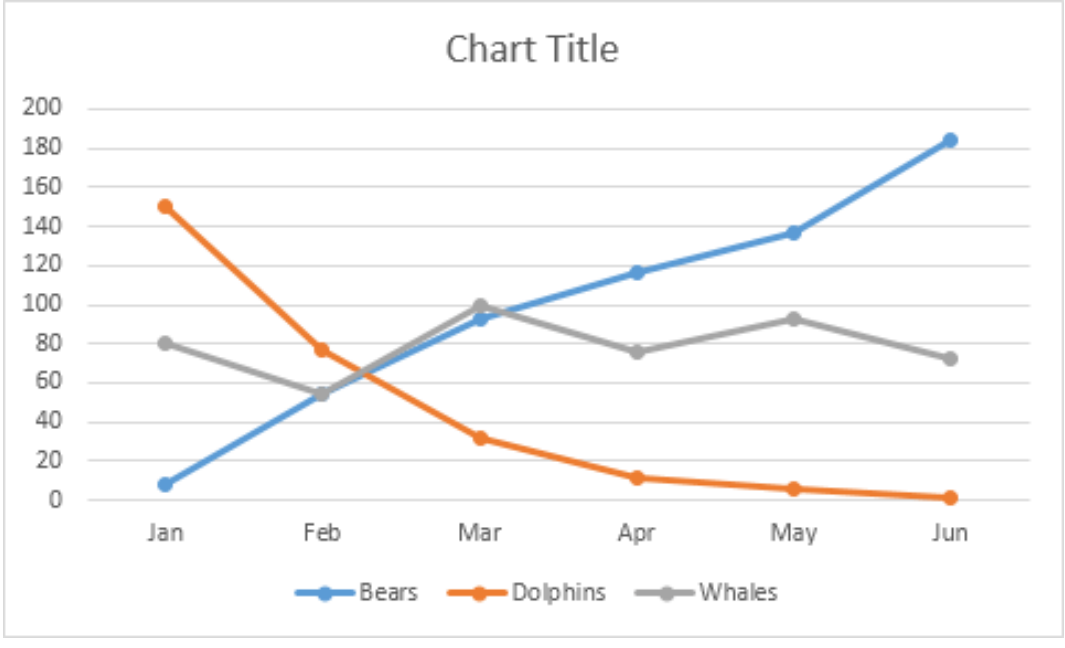




Excel Charts And Graphs

To create a chart in Excel:

3. Customize your chart's design, layout, and formatting to meet your requirements.



Exercise 1. Data on a single variable.

Exercise 1. Data on a single variable.

- Data set: `ages.xlsx`
- The (imaginary) ages of survey respondents, $n = 30$.

Part 1. Summary statistics.

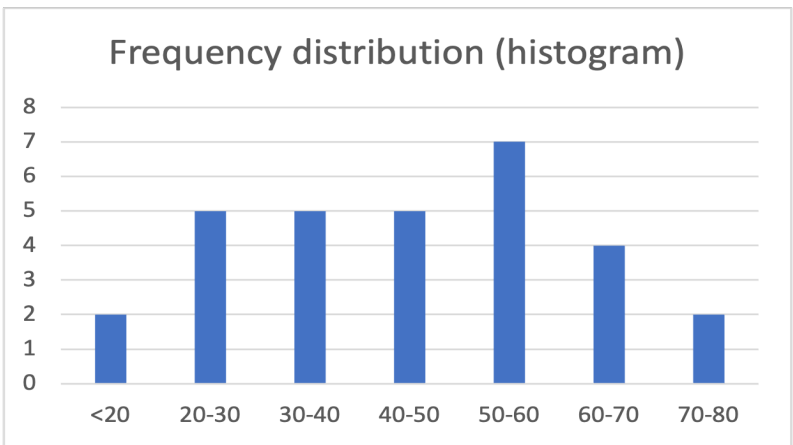
1. Create a new tab on the spreadsheet.
2. Compute the mean, median, min, max of ages using Excel functions. Make a table.
3. Compute the mean age using only the following excel commands: SUM(), COUNT().
4. Which one is higher, the mean or the median? What does this tell us about the shape of the distribution of the data?

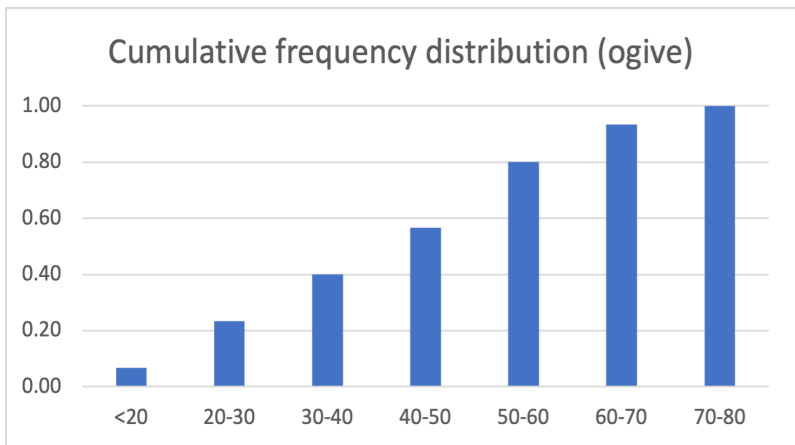
Part 2. Plotting data.

1. Create a new tab on the spreadsheet with the data from the original tab.
2. Compute a frequency distribution table. Decide yourself the cutoff points.
3. Compute a corresponding cumulative distribution table.
4. Make a graph describing the frequency distribution. The title should be *"Frequency distribution (histogram)"*.
5. Make a graph describing the cumulative frequency distribution. The title should be *"Cumulative Frequency distribution (ogive)"*.

Part 2. Plotting data.

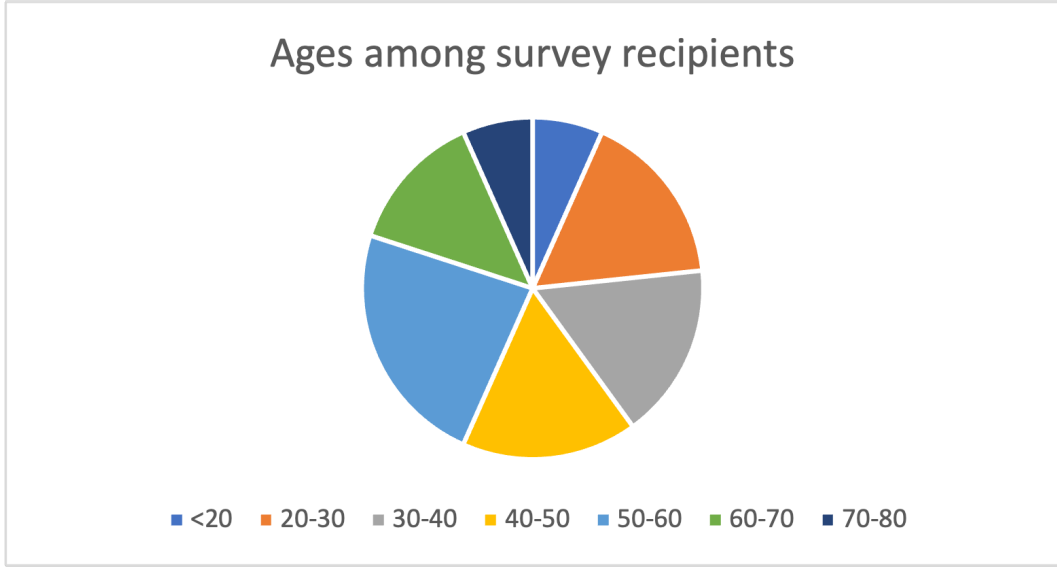
- Two bar charts describe *frequency distribution* and *cumulative frequency distribution*.





Part 3. Plotting data: Pie charts.

- A pie chart summarizes the age distribution.



- It emphasises the proportions of frequencies in each age group.

Exercise 2. Data on multiple variables.

Exercise 2. Data on multiple variables.

- Data set: `incomes.xlsx`
- About (imaginary) data on two variables
  - income (`inc`), and
  - years of schooling (`educ`), which is categorical and coded as follows:

9	Grade 9
10	Grade 10
11	Grade 11
12	Grade 12
13	1 year of college
...	
17	5 years of college
18	6+ years of college

Part 1. Correlation

- Open up the data and create a new tab with the original data.
- Compute the sample mean and the sample variance of both variables.
- Compute the coefficient of correlation between income and years of schooling. How do we interpret the correlation coefficient? Is it a large or a small coefficient?

Part 1. Correlation

- Coefficient of Correlation summarises the *direction* and *strength* of the linear relationship between two quantitative variables.
- Sample correlation coefficient

$$r = \frac{s_{xy}}{s_x s_y}$$

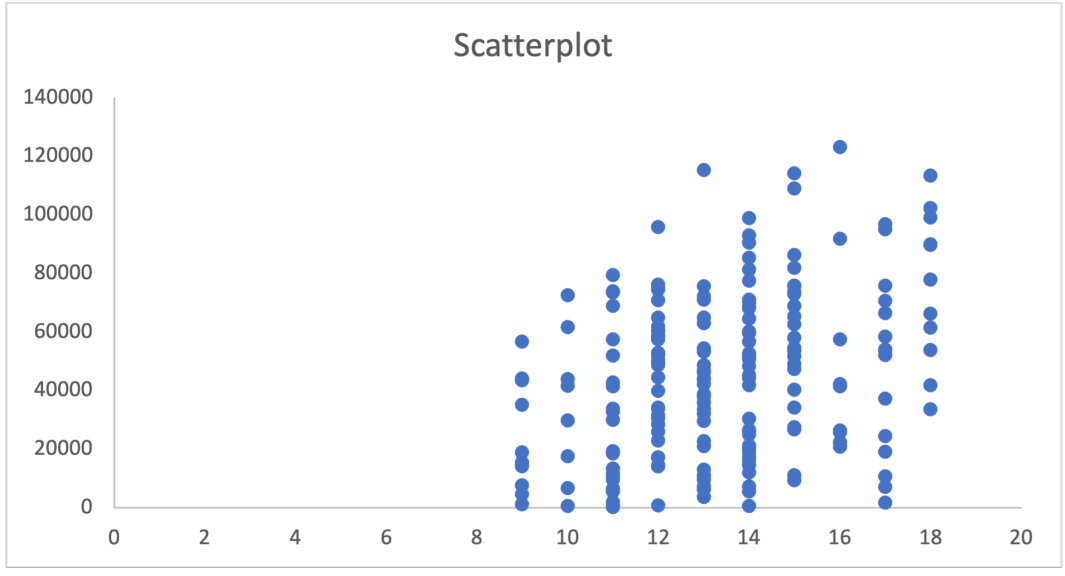
- **Magnitude of correlation**
  Often, when we discuss correlations, we use the following rules of thumb:
  - $|\, r \,| < 0.2$: small correlation
  - $0.2 \leq |\, r \,| \leq 0.8$: medium correlation
  - $|\, r \,| > 0.8$: very strong correlation

Part 2. Scatterplots

- Create a new tab with the original data.
- Make a scatterplot of the data. This is a plot where each individual (each row in the data) is described as a dot, and the x-axis value shows the years of education, the y-axis value shows the income of the individual.
- How does the scatterplot reflect the correlation coefficient from Part 1?

Part 2. Scatterplots

- A scatterplot of the data where each individual (each row in the data) is described as a dot, and the x-axis value shows the years of education, the y-axis value shows the income of the individual.

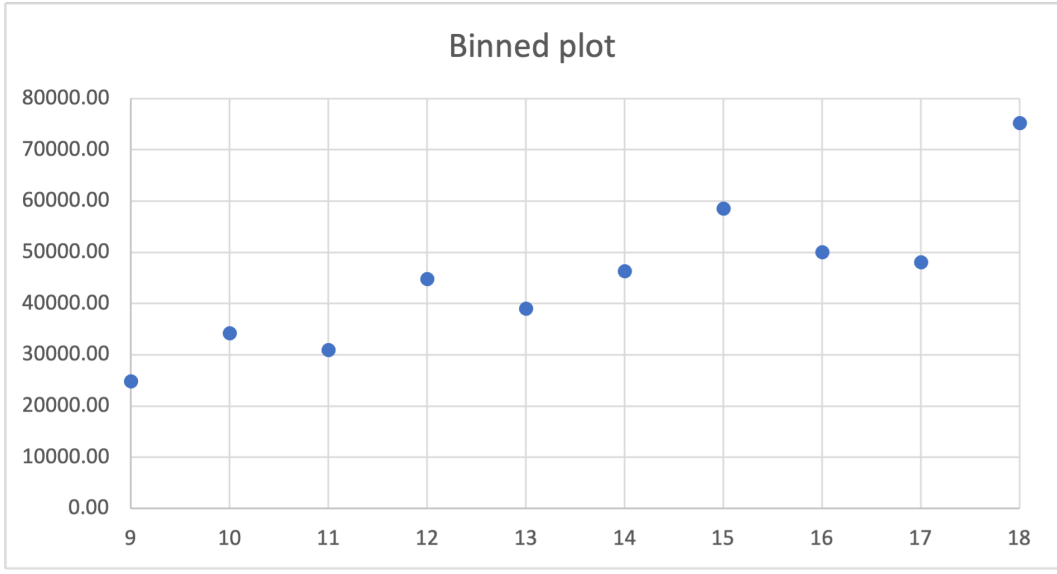


Part 3. Binned plots

- The scatterplot can be hard to read when there are many observations. Different data visualization tools can make the data easier to interpret.
- Create a new tab with the original data.
- For each value of "educ", compute the conditional average of income. That means that for each level of education you calculate the average income
  (Hint: Use Excel formulas UNIQUE() and AVERAGEIF()).
- Make a binned plot where x-axis is the years of education and y-axis is the average income for the given level of education.
  (Hint: There will be only one dot in the picture for each years of schooling).
- When can a binned plot be more informative than a scatterplot?

Part 3. Binned plots

- A binned plot where x-axis is the years of education and y-axis is the average income for the given level of education.



- There will be only one dot in the picture for each years of schooling.

Part 3. Binned plots

- A binned plot is often used when we want to describe the association (correlation) between two variables, but there are too many observations to make an informative scatterplot.

Part 4 [opt]. Percentiles and IF-statements on Excel

- Compute the 50^{th} percentile of the data (the median) conditional on years of education.
  (Hint: you should calculate the 50^{th} percentile for all values of years of education; use formula MEDIAN(IF()))
- Compute the 25^{th} and the 75^{th} percentiles of the data conditional on years of education.
  (Hint: use formula PERCENTILE(IF()))
- How do we interpret the table?

REFERENCES

1. Excel Cheetsheet: https://blog.enterprisedna.co/excel-cheat-sheet/
2. Create charts in Excel: https://www.excel-easy.com/data-analysis/charts.html#google_vignette