Chapter 1

Exploratory Analysis

Statistics is a science of data. To study statistics, we have to describe data from some proper perspectives. Generally speaking, there are two ways to describe data, graphical description and numerical description. Those are what we are going to learn in this section.

1.1 Basic concepts

Table 1.1: Final scores

Names	Student.id	Calculus	Physics	Gender
James	170101	75	74	Male
Sam	170102	87	83	Male
Crystal	170103	88	92	Female
Evelyne	170104	84	82	Female
Phoebe	170105	89	86	Female
Vince	170106	76	85	Male
Mike	170107	71	76	Male
Allen	170108	73	60	Male
Lucy	170109	81	89	Female
Kitty	170110	86	91	Female
Owen	170111	88	82	Male
Angela	170112	96	93	Female
Christina	170113	87	81	Female
Jamie	170114	84	80	Female
Meggie	170115	80	82	Female
Kevin	170116	97	99	Male
Tom	170117	90	86	Male
Lunna	170118	85	62	Female
John	170119	81	88	Male
Jason	170120	72	80	Male

- *Table 1.1* gives the final scores of 20 students, each student is an **individual**.
- All stduents are described through *Names, Students id, Calculus, Physics*, and **Gender**. They are called **Variables**, for they may take different values for different individual.s
- The values of *Calculus* and *Physics* can be manipulated as usual numbers. *Calculus* and *Physics* are called **quantitative variables**.
- The values of *Names, Students id* and *Gender* can not be manipulated as usual numbers. They are called **categorical variables**.
- The way a variable takes different values is call the **distribution** of this variable.

1.2 Graphical displays

- **Bar Graph**, a graph to show the distribution of categorical variables .

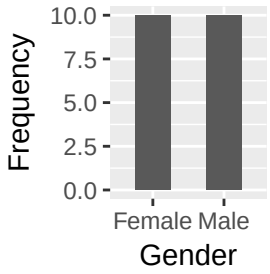

Figure 1.1: Bar graph for the distribution of the Gender

Be sure to indicate the labels of the aixes whenever a graph is drawn

The vertical axis of *figure* 1.1 is **frequency**, which means the number of individuals. *Figure* 1.1 means there are 10 male students and 10 female students.

Sometimes the vertical axis can be **relative frequency**, as shown in *figure*1.2. It means 50% of the students are male and 50% are female.

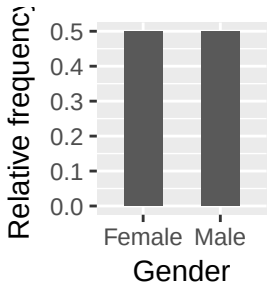

Figure 1.2: Bar graph for the distribution of the Gender

- **Dot plot**

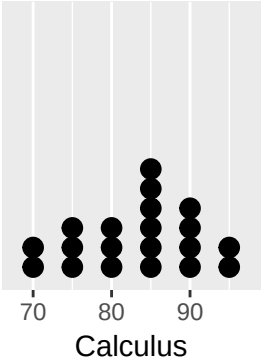

Figure 1.3: Dot plot for the distribution of the Calculus scores

- **Histogram**

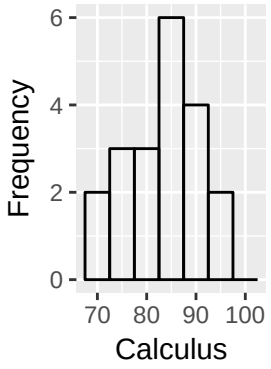

Figure 1.4: Histogram for the distribution of the Calculus scores

What is the difference between the bar graph and histogram?

- **Stemplots**

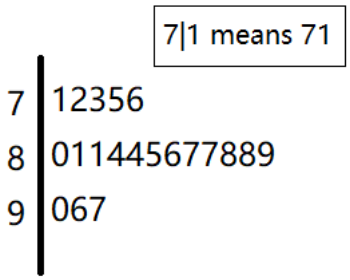

Figure 1.5: Stem plot for the distribution of calculus scores

Don't forget the legend