Descriptive Statistics, pt. I

ECON 3640–001

Marcio Santetti
Spring 2022

Motivation

Getting to know our data

Before any complex operations, we need to **know our data**.

Such previous assessment may **not** appear in a final work (e.g., a *report*, a *paper*, a *project*), but is the **first step** of any data analysis procedure.

Some key statistical concepts

Key statistical concepts

Two of the *most important* statistical concepts:

- **Population**: a group of *all* items of interest to a statistics practitioner.
  - A descriptive measure of a population is called a *parameter*.

- **Sample**: A *set* of data drawn from the studied population.
  - A descriptive measure of a sample is called a *statistic*.
  - It is used to make *inferences* about population parameters.

Types of data and information

Types of data and information

In Statistics, a **variable** is a characteristic of a population or sample.

3 main **types** of data in Statistics:

- *Interval* data;
- *Nominal* data;
- *Ordinal* data.

Types of data and information

Interval data is also known as *quantitative* or *numerical* data.

- Examples?

Nominal data comprehends *qualitative* or *categorical* variables.

- Ordering *does not* matter.
- Examples?

For **ordinal data**, ordering *does* matter.

- Examples?

Graphical statistical techniques

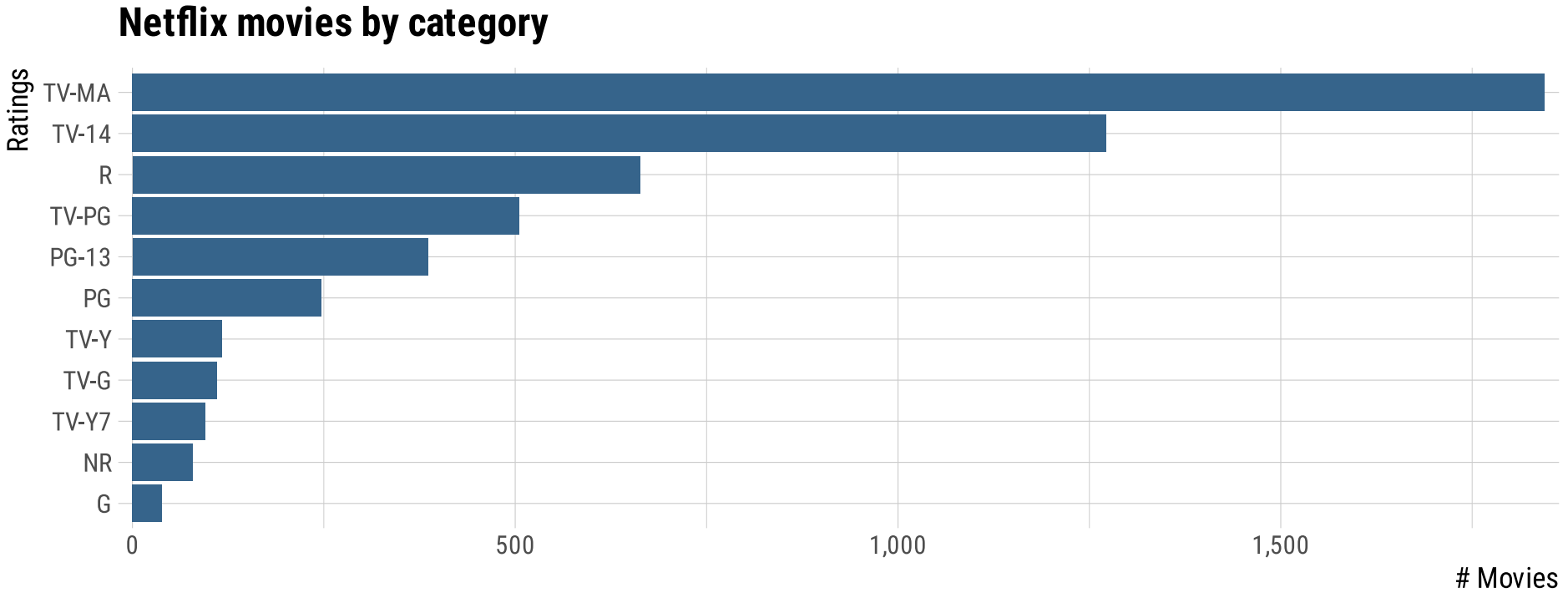
Graphical statistical techniques

Sometimes, a *table* describing nominal (or ordinal) data may not be the best option to catch a reader's eye.

Therefore, some graphical techniques can be useful, such as the **bar** and the **pie** charts.

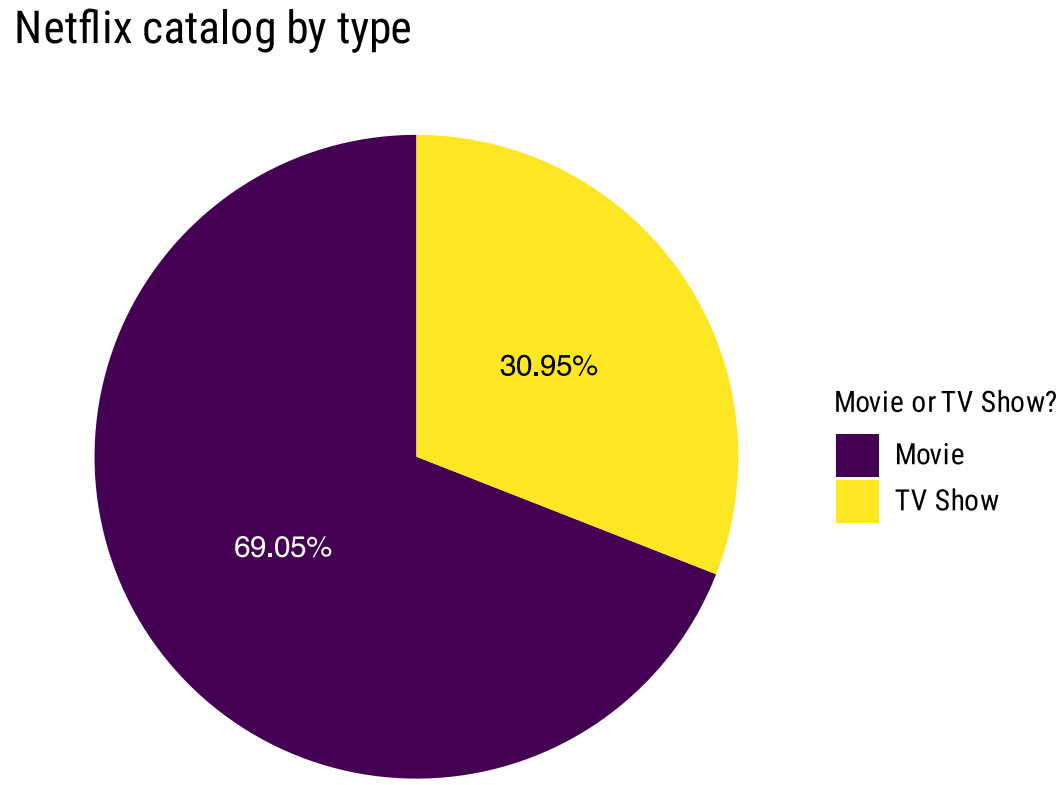
Graphical statistical techniques

A **bar chart** is useful to illustrate an *absolute frequency*, i.e., the total number of observations fitting a given category in a data set.



Graphical statistical techniques

If, on the other hand, we want to illustrate *relative frequencies*, i.e., the percentage with which each category appears in the data set, a **pie chart** is the best option.



Graphical statistical techniques

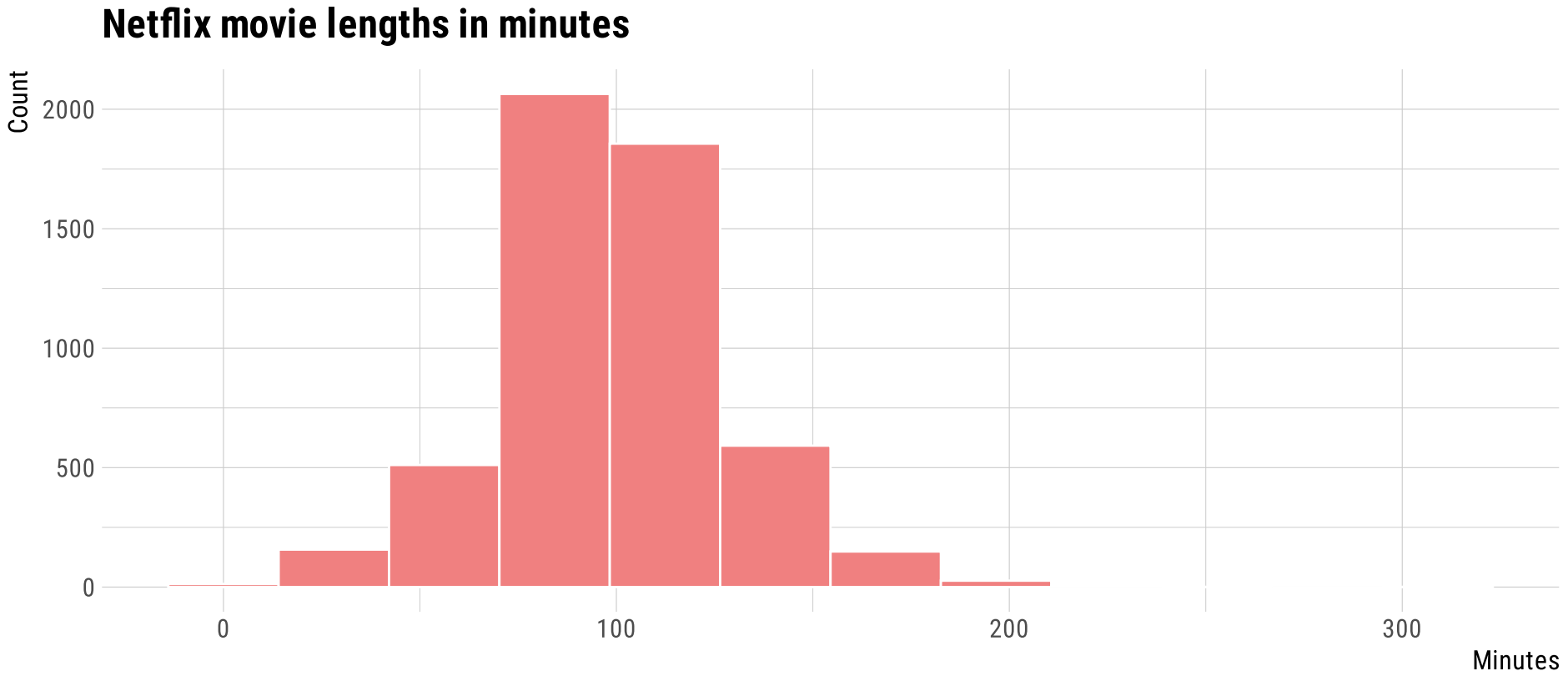
Moving on to **interval data**, one of the most common ways of presenting it is through a **histogram**.

To construct a histogram, the easiest way to start is by following a *recipe*:

1. Find the data set's *lowest* and the *highest* values;
2. Define the appropriate intervals (*bin size*), and the number of *observations* contained in each interval;
3. Draw *each* bin next to each other.

Graphical statistical techniques

An example:



Graphical statistical techniques

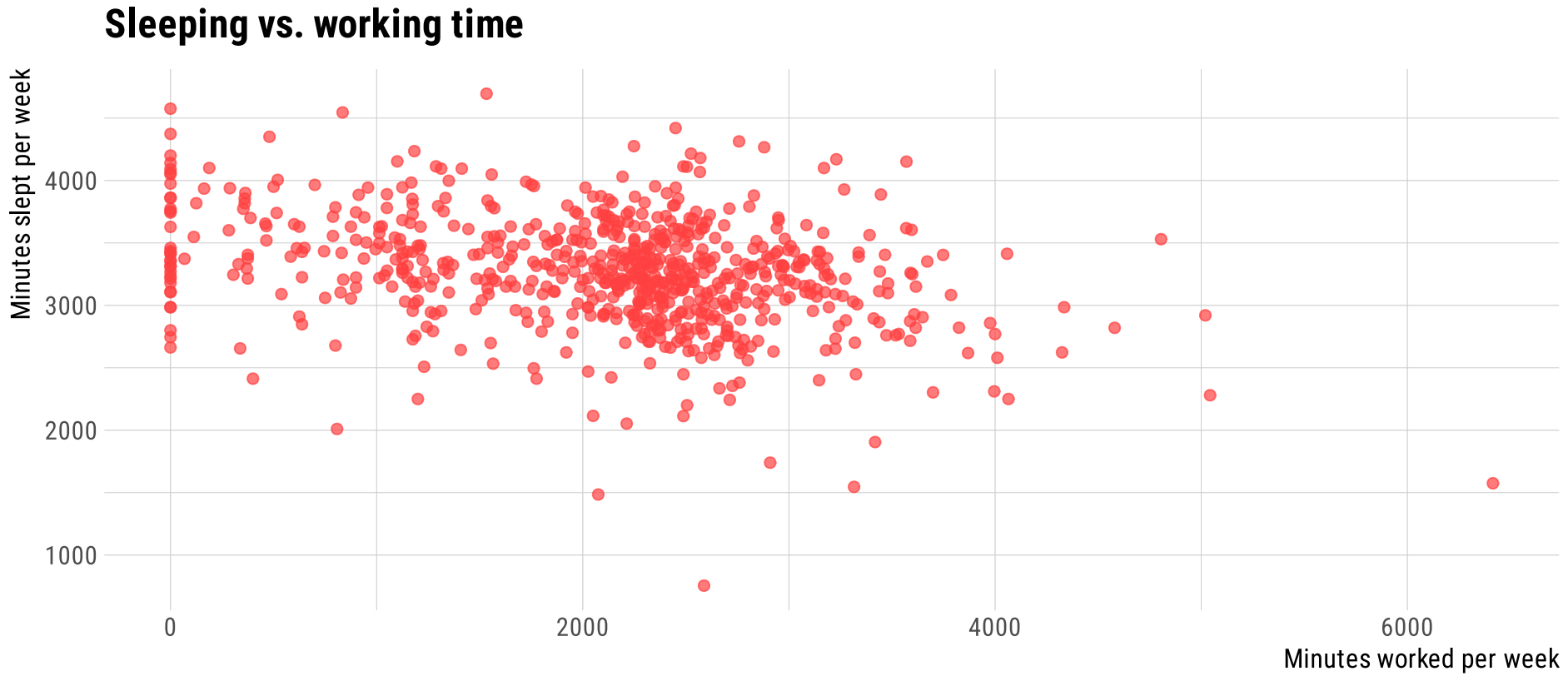
Suppose that we want to visually describe the relationship between **two** interval variables.

The technique to be used is the **scatter diagram**.

To draw a scatter plot, we must define an *independent* (explanatory) and a *dependent* (explained) variable.

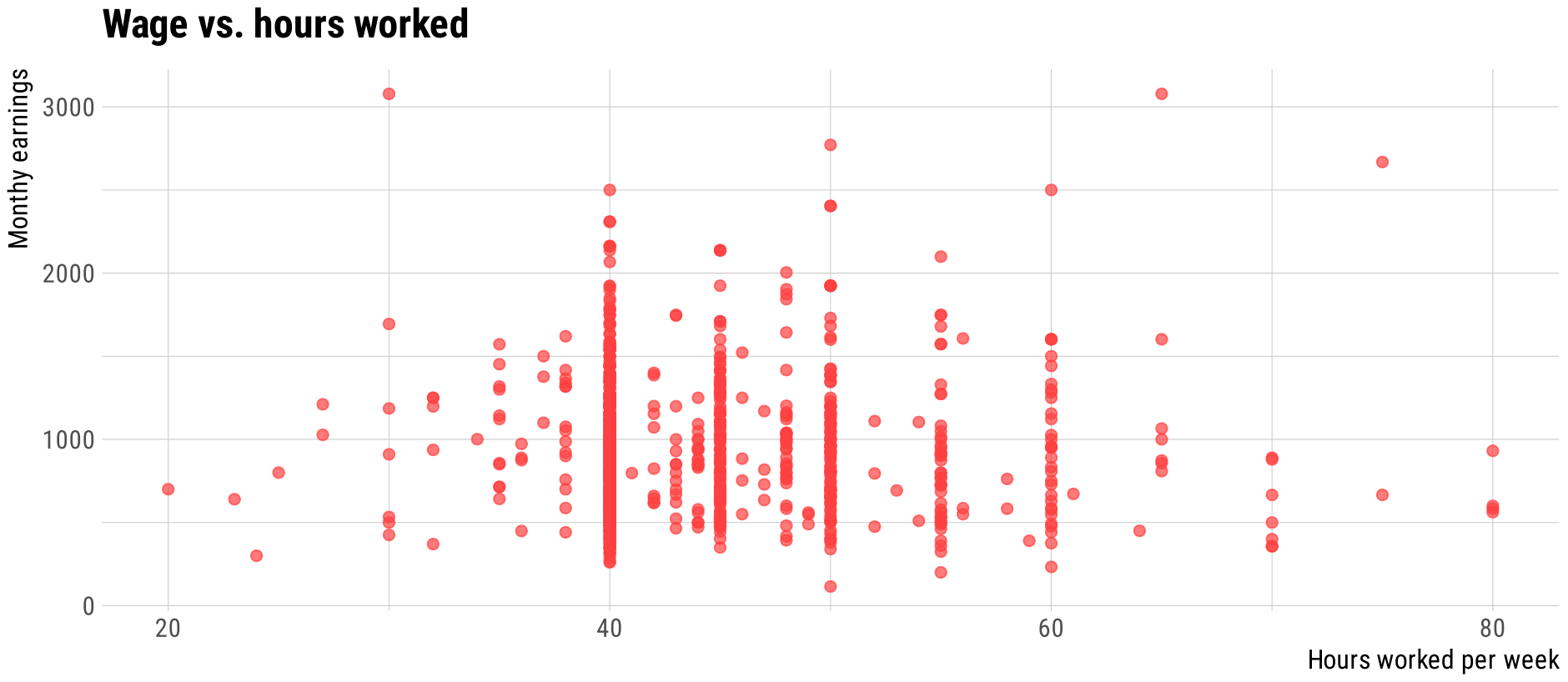
Graphical statistical techniques

Example 1:



Graphical statistical techniques

Example 2:



Next time: Statistical measures