Terminology in Statistics

Niall McGuinness

2022-11-30

# What is markdown?

## Heading 2

## Heading 3

This is in **bold**

* Player - data
* Customer - data

(Dkit)[“www.dkit.ie”]

| Table Header | Second Header |
| --- | --- |
| Table Cell | Cell 2 |
| Cell 3 | Cell 4 |

# What are the key terms in statistics?

## What types of variables exist in a game-related survey?

* Name - character - “John”
* Gender - character - m/f/o/u
* Age - integer - 45, 32
* Satisfied - integer - 1->5, unsatisfied/satisfied
* Time - float-point - 2.25hrs
* Price - float-point - 2.25hrs

### Define categories for the examples above

* Numerical - continuous, discrete e.g. height, age
* Categorical - nominal (name) e.g. “brown”, “male”, “satisfied”

### What is a statistic?

* If we derive a value/variable from a sample we call it a statistic e.g. average age of a 100 people then we have a **sample statistic**.
* If we derive a value/variable from the population we call it a parameter e.g. census gives us average age of the population then we have a **population parameter**.

statistics -> samples parameters -> population

### What is an example of a statistic?

## [1] 3.2 2.0 5.1 1.0 4.0 5.0 3.0 2.2 1.0 4.0 5.1 1.0 2.8 3.3 1.0

## [1] 2.913333

## [1] 1

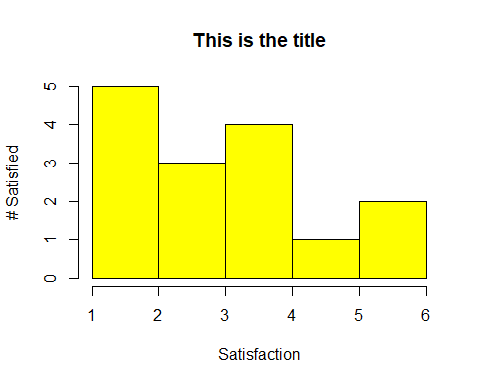
## [1] 3.5

This average of the data is 2.91 satisfaction points.

## What can we do with R Studio?

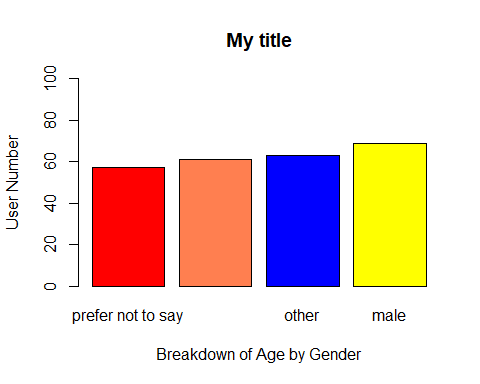
Well, we can generate plots really quickly from CSV file, or from remote RBDMS data, or from CSV file on web page.

### Histogram



### Barplot

gamer\_survey\_data <- read.csv("amalgamated\_game\_survey\_250\_2022.csv")  
barplot(sort(table(gamer\_survey\_data$gender), decreasing = FALSE),  
 main = "My title",  
 xlab = "Breakdown of Age by Gender",  
 ylab = "User Number",  
 ylim = c(0, 100),  
 col=c("red", "coral", "blue", "yellow"))



### Boxplot

### QQ Plot