Statistics Tutorial

Day 1

Prabesh Dhakal

2020 April 09

Who am I?

PRABESH DHAKAL

DATA SCIENCE MASTER

WHAT I DO

- About to write Master Thesis in data science.
- Work in Henrik's lab and support scientists with their data related challenges:
  - data collection, preparation, analysis, and communication

WHAT I USE

R

Python

What are we doing today?

1. Introduce the course

What topics are we going to cover during the future weeks in this semester?

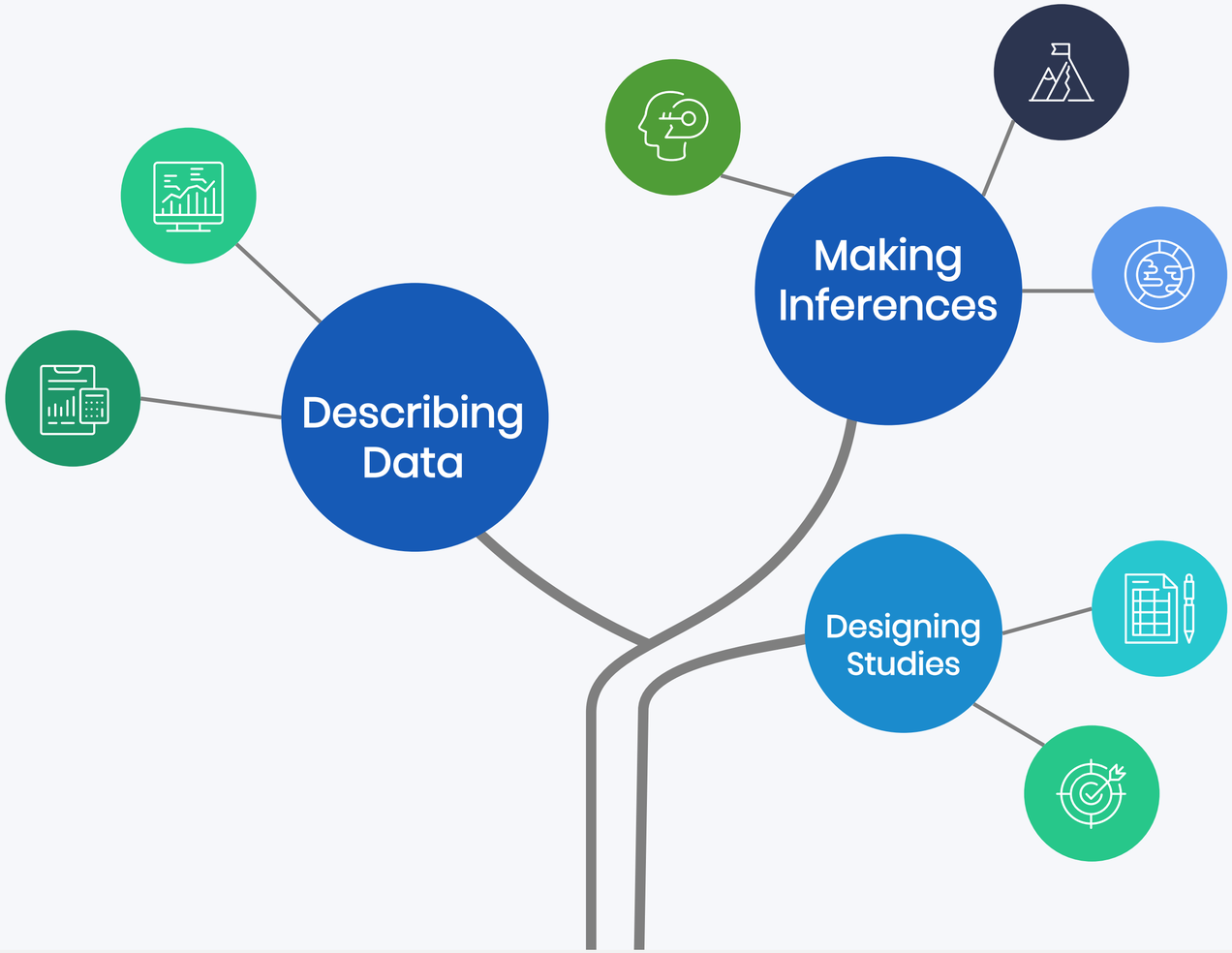
2. Discuss some tools

Setup the tools required for the module on the devices. Familiarize ourselves with the basic setup of the tools.

3. Answer your questions

Discuss a few basic things to get a gentle start to the tutorials.

What will we cover in the future?



We will rely on R and RStudio to understand and use most of these topics.

What will we cover in the future?

(For those who prefer a list)

- Data types
- Distribution of data
- Probability and Probability Distribution
- Simple tests
- Correlation
- Regression
- Analysis of variance (ANOVA)
- Types of experiments …

Discuss Tools

- Setup the tools required for the module on the devices.
- Familiarize ourselves with the basic setup of the tools.

Install Softwares

You start using R-Studio by doing the following things:

Step 1: Install R

Step 2: Install R-Studio

Windows Users

Download Links:

- **Download R** (Windows)
- **Download RStudio** (Windows)

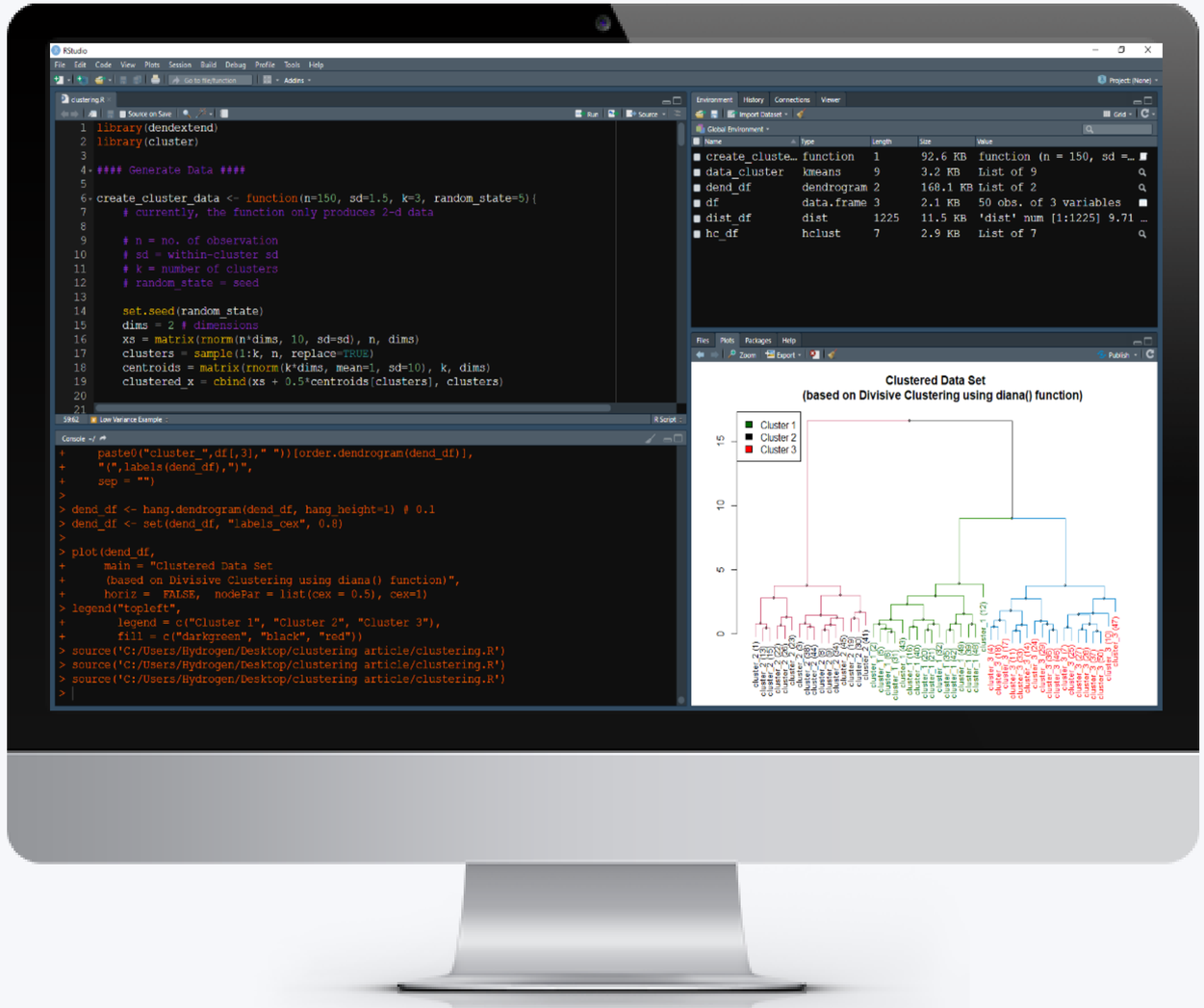
Mac Users

Download Links:

- **Download R** (Mac)
- **Download RStudio** (Mac)

Linux Users

You know what to do.



PROGRAMMING PARADIGM: R

WHAT IS R?

Short version:

a calculator on steroids

A slightly longer explanation:

R is a programming language **purpose built for data analysis**

- easy to learn,
- works in all desktop environments, and
- has a great community of users and supporters.

MAIN TOOL WE WILL USE: RStudio®

Prepackaged Tools

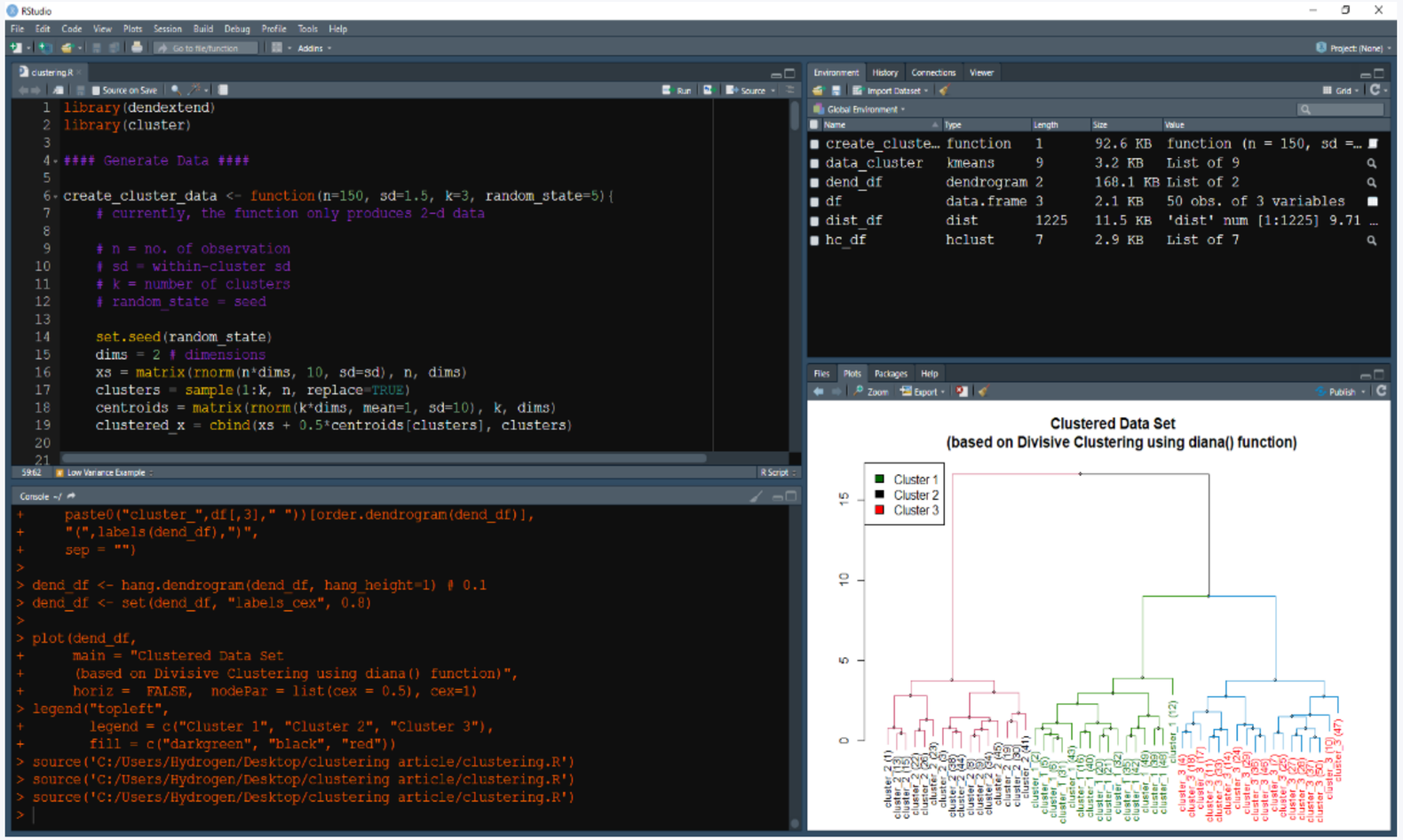
Its graphical interface makes working with R even easier.

A better alternative to Excel/SPSS/. . .

Lighter, faster, better support, better community engagement …

Free and Open Source

Its not a proprietary software. The program is community driven.



A walkthrough of **R** and **R-Studio**

One more note.

Moving forward, when you see  symbol on a slide (usually on the top left side), you are encouraged to try solving the problem on the screen.

Questions so far?

Tutorial & Discussion

1. Start of the tutorial.
2. General Q&A

I want to hear your thoughts.

(Mentimeter Quiz)

STATISTICS

What is statistics?

Statistics is the science of collecting, preparing, analyzing, and interpreting data.

2 key parts:

- **Collecting and Preparing Data**

  Involves performing experiments and/or surveys, collecting data, and storing it in an organized format.

- **Analyzing Data and Making Inferences**

  Involves taking the data, summarizing it, and performing tests.

STATISTICS: WHY?

Why should we care about statistics?

Video
Can Maths Predict the Future? - Hannah Fry

Statistics is really one of the essential tools that we can use to make sense of the world around us.

It allows us to make discoveries, make decisions, and predict events.

SUMMING UP

1. Statistics is the science of collecting, preparing, analyzing, and interpreting data.
2. Statistics allows us to:
   a) make sense of the world around us,
   b) make decisions, and
   c) make predictions.
3. This course will complement the lectures and we will apply the concepts discussed there with real data.
4. The tools we will use to work with the data and perform calculations is going to be R-Studio.

That’s it for today!

Next week, we are going to discuss types of data.

Tasks for you:

- Install R and RStudio on your devices.

If you want to reach me, mail me at:

`prabesh.dhakal@stud.leuphana.de`