

# Welcome!

Machine Learning with R  
Basel R Bootcamp



October 2019

# Welcome to the Basel R Bootcamp!

Machine Learning with R  
October, 2019

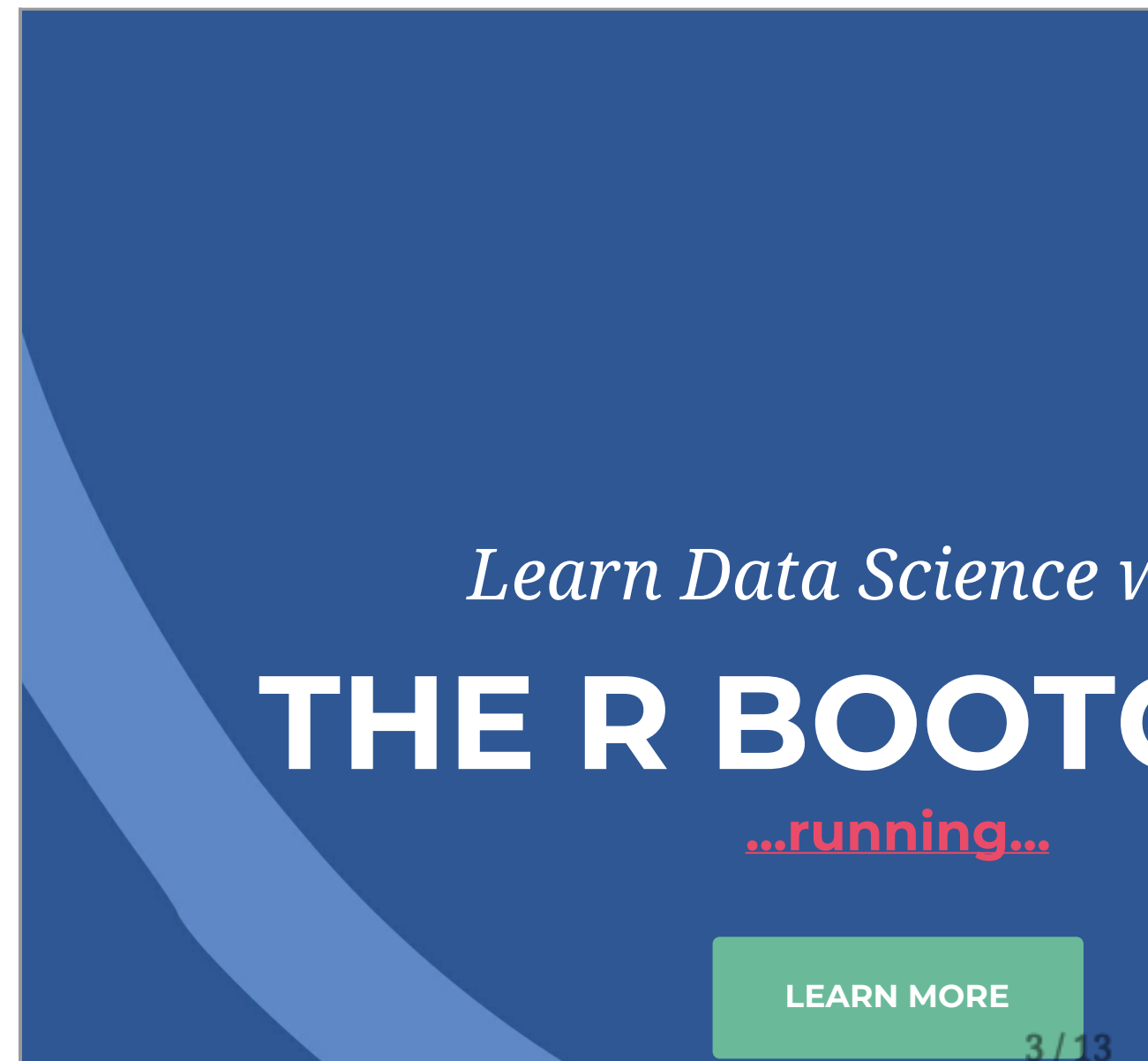


from [swisskyline.ch](http://swisskyline.ch)

## Schedule and Materials

Click the "...running..." link on our homepage!

[www.therbootcamp.com](http://www.therbootcamp.com)



# Aim

The aim of this workshop is enable you to thoughtfully implement machine learning methods, using R, to make data-driven decisions.



# What we will cover

1. Intro to the **fundamentals** of ML.
2. Intro to a few influential ML **models**.
3. Intro to **using R** for ML.
4. Opportunity to **learn by doing**.
5. Opportunity for **discussion**.

## What we will *NOT* cover

In-depth, advanced treatment of ML.



from [machinelearningplus.com](http://machinelearningplus.com)

# Schedule

We have an **ambitious schedule!**

2 days, 3-4 sessions per day, ~2.5 hours per session, on a variety of topics.

Each session contains a slide-based **introduction** and a hands-on **programming interactive or practical**

	Sat, 5 Oct	Sun, 6 Oct
0900	Welcome	Recap
0930	<b>What is ML</b>	<b>Optimization</b> +Practical
1045	<b>R for ML</b> +Interactive	
1200	Lunch	Lunch
1300	<b>Fitting</b> +Practical	<b>Features</b> +Practical
1500		<b>Models</b> +Practical
1530	<b>Prediction</b> +Practical	
1715		<b>Look ahead</b>
1800	Wrapup	Apero

# Introduction

**30 - 45 minutes** slide based introduction to help you understand the basic concepts, including:

1. Theory
2. Code examples
3. Live demonstrations

**All available online.** Follow along, and don't worry about memorizing.

**This is a Link to our materials.**

Be interactive! **Ask questions!**



from [Freepik.com](https://www.freepik.com)



from [Freepik.com](https://www.freepik.com)



# Practicals

**Write your own** R scripts!

Anywhere from 10 ~ 30 programming tasks. Starts easy, gets **progressively more challenging**.

Go at your **own pace** and finish as much as you can. We have a **diverse group** with different backgrounds and interests, we love that!

We'll give you all the **answers later**.

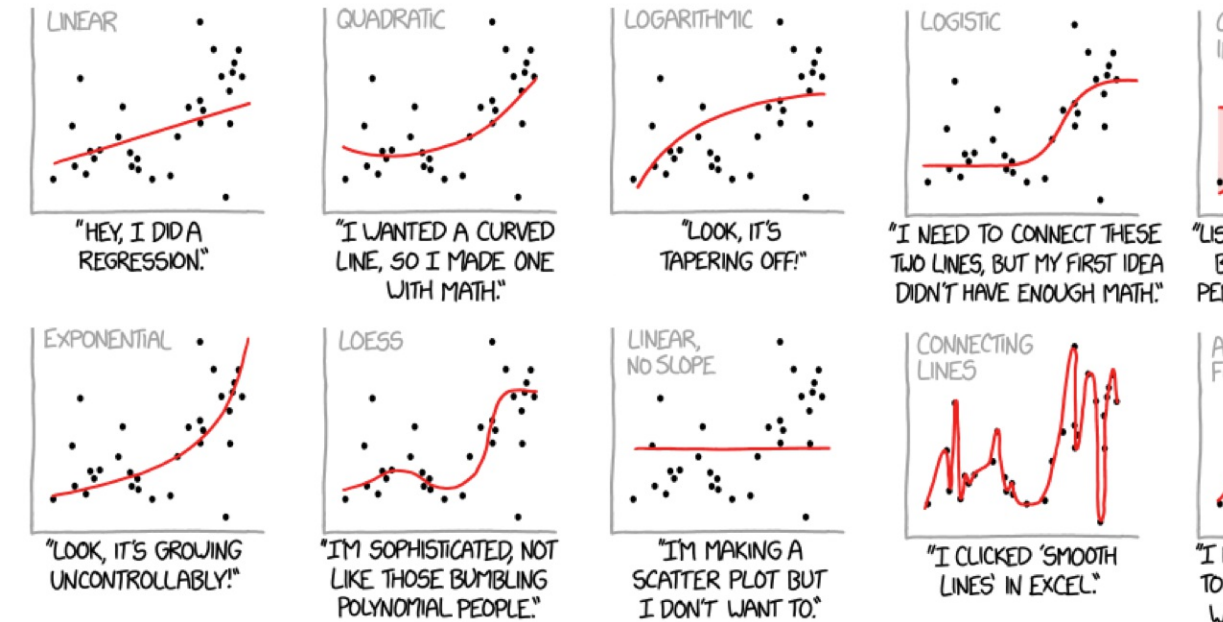
Be interactive! **Ask questions!**

## Fitting

Machine Learning with R

🔔 ([https://therbootcamp.github.io/ML\\_2019Oct/](https://therbootcamp.github.io/ML_2019Oct/)) 🏠 (<https://therbootc>)

### CURVE-FITTING METHODS AND THE MESSAGES THEY SEND





# 1-1 Coaching

Learning a computer language is **hard**.

**Take breaks when you need them**, and don't get discouraged!

Sometimes, you may find yourself in need of some **intensive help**. That's ok! We're here to help!

At any time, you can **request a 15 minute 1-1 intensive** coaching session. We'll grab a coffee and work through it together.

*WHEN YOU HEAR THIS:*



from [geek-and-poke.com](http://geek-and-poke.com)

# Cheatsheets

## Base R

Base R Cheat Sheet	
<b>Getting Help</b>	
Accessing the help files	
?function: Get the help file for the function named function.	
help: Get the help file for the function named function.	
help.search: Search for the function named function.	
help.start: Open the help browser.	
Using Libraries	
library: Load a package into the current session.	
require: Load a package into the current session, with an error if it is not installed.	
install.packages: Install a package from CRAN.	
update.packages: Update all installed packages.	
Working Directory	
getwd: Get the current working directory.	
setwd: Set the current working directory.	
Use projects to manage the working directory for the whole project.	

## Data Import

Data Import :: CHEAT SHEET	
<b>Read Tabular Data</b>	
read.csv: Read a CSV file into a data frame.	
read.csv2: Read a CSV file into a data frame, using comma as the delimiter.	
read.delim: Read a delimited file into a data frame.	
read.table: Read a table from a file into a data frame.	
read.fwf: Read a fixed-width formatted file into a data frame.	
readLines: Read a text file into a character vector.	
write.csv: Write a data frame to a CSV file.	
write.csv2: Write a data frame to a CSV file, using comma as the delimiter.	
write.delim: Write a data frame to a delimited file.	
write.table: Write a data frame to a table file.	
write.fwf: Write a data frame to a fixed-width formatted file.	
writeLines: Write a character vector to a text file.	
<b>Save Data</b>	
save: Save an R object to a file.	
load: Load an R object from a file.	
saveRDS: Save an R object to a file in RDS format.	
loadRDS: Load an R object from a file in RDS format.	
<b>Read Non-Tabular Data</b>	
read.json: Read a JSON file into a list.	
read_xml: Read an XML file into an R6 object.	
read_yaml: Read a YAML file into a list.	
read_gg: Read a ggplot2 plot into a ggplot object.	

## Data Transformation

Data Transformation with dplyr :: CHEAT SHEET	
<b>Filter</b>	
filter: Filter rows based on one or more conditions.	
select: Select columns by name or index.	
rename: Rename columns.	
arrange: Arrange rows in ascending or descending order.	
distinct: Get unique values of a column.	
sample_n: Get a random sample of rows.	
sample_frac: Get a random sample of rows, by fraction.	
<b>Summarise</b>	
summarise: Summarise data into a single row.	
count: Count the number of rows in each group.	
pull: Extract a column from a data frame.	
do: Execute an R expression for each group.	
<b>Group</b>	
group_by: Group data by one or more variables.	
ungroup: Ungroup data.	
group_vars: Get the names of the grouping variables.	
n: Get the number of rows in each group.	
first: Get the first row in each group.	
last: Get the last row in each group.	
nth: Get the nth row in each group.	
nth_value: Get the nth value in each group.	
lead: Get the next row in each group.	
lag: Get the previous row in each group.	
<b>Join</b>	
inner_join: Join two data frames on common columns.	
left_join: Join two data frames, keeping all rows from the left table.	
right_join: Join two data frames, keeping all rows from the right table.	
full_join: Join two data frames, keeping all rows from both tables.	
semi_join: Join two data frames, keeping only rows from the left table that have matches in the right table.	
anti_join: Join two data frames, keeping only rows from the left table that do not have matches in the right table.	

## Caret

caret Package	
<b>Specifying the Model</b>	
train: Train a model on a dataset.	
trainControl: Specify the resampling method and other options.	
modelObject: Create a model object.	
<b>Preprocessing</b>	
preProcess: Preprocess a dataset.	
trainPreProcess: Preprocess a training dataset.	
testPreProcess: Preprocess a test dataset.	
<b>Performance Metrics</b>	
resample: Resample a dataset.	
getMetrics: Get the performance metrics for a model.	
<b>Grid Search</b>	
trainGrid: Train a model on a grid of parameter values.	
getBestTune: Get the best parameter values.	
<b>Adding Options</b>	
addOption: Add an option to a model.	
<b>Resampling Options</b>	
resampleMethod: Specify the resampling method.	
resampleSize: Specify the size of the resampled data.	
<b>Parallel Processing</b>	
makeParallel: Create a parallel computing environment.	
trainParallel: Train a model in parallel.	

# Breaks

Programming is exhausting! **Take breaks** when you need them!

Help yourself to **refreshments** and as much coffee as you can drink.

Lunch on your own or **walk with the group** to **Tibits** or the **Union Diner**

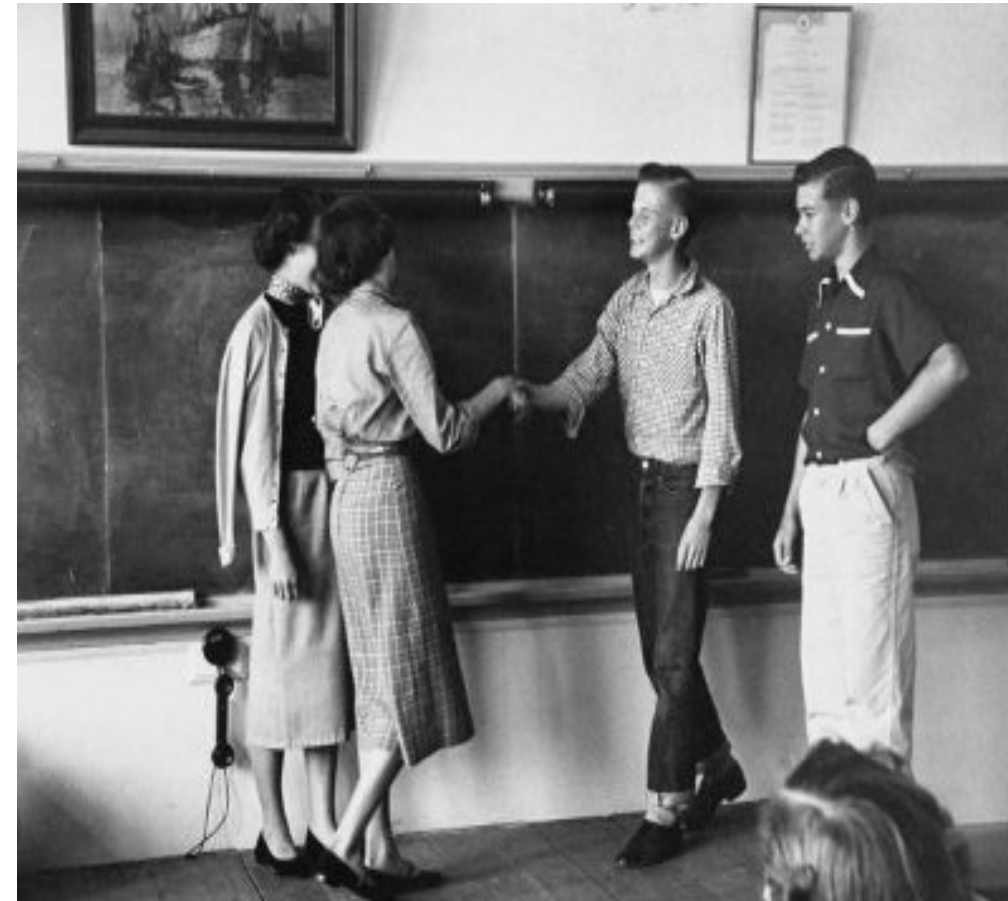
Be back in the room at **13:00 sharp** so we can continue learning!



from [Tibits.ch](https://tibits.ch)

# Introductions

1. What is your name?
2. Where are you from?
3. What is your current profession?
4. What is your programming experience with R and in general?
5. What is your experience with ML?
6. What would you most like to do with ML?
7. Coffee or tea?
8. Beer or wine?
9. Berlin or Paris?



from [artofmanliness.com](http://artofmanliness.com)

# Schedule