R-Ladies

from scratch

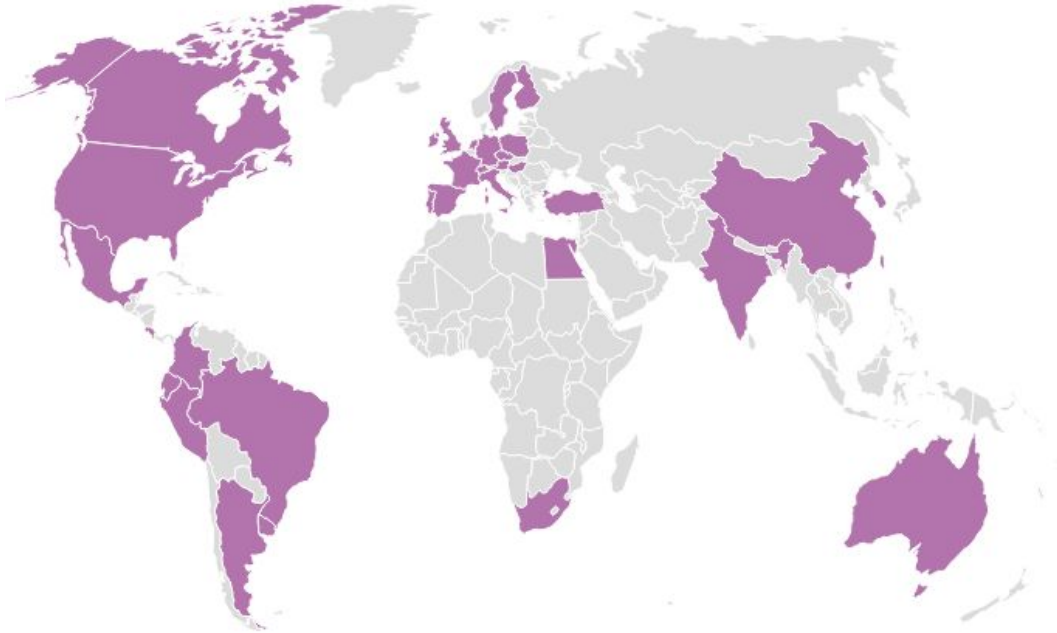
Kasia Kulma & Erle Holgersen

OVERVIEW

1. R-ladies - who are we? (5 mins)
2. What's R and why you should learn it (5 mins)
3. Let's learn some R! (1h 20 mins)
4. Post-session survey (5 mins)

What's R-Ladies?

R-ladies is a world-wide organization to promote gender diversity in the R community: https://rladies.org/



- 60+ Chapters worldwide
- 20+ Countries
- 6000+ Members

What's R...

- Open source programming language for statistical computing & graphics
- Originated from language S
- Developed & matured in early 1990's in New Zealand (Auckland University)
- Names to remember: Ross Ihaka & Robert Gentleman

...and why you should learn it

- Freeware, open - source & platform - independent
- **HUGE** choice of analytics and statistics libraries
- Integrates many data - sources
- Efficient data structures (data.frame, tibble, data.table, etc.)
- Produces arguably best Data Visualizations
- Can build interactive web apps (Shiny)
- **Easy** reporting and blogging (RMarkdown, blogdown)
- **AWESOME** community

...and why you should learn it

- Freeware, open - source & platform - independent
- **HUGE** choice of analytics and statistics libraries
- Integrates many data - sources
- Efficient data structures (data.frame, tibble, data.table, etc.)
- Produces arguably best Data Visualizations
- Can build interactive web apps (Shiny)
- **Easy** reporting and blogging (RMarkdown, blogdown)
- **AWESOME** community

...and why you should learn it

- Freeware, open - source & platform - independent
- **HUGE** choice of analytics and statistics libraries
- Integrates many data - sources
- Efficient data structures (data.frame, tibble, data.table, etc.)
- Produces arguably best Data Visualizations
- Can build interactive web apps (Shiny)
- **Easy** reporting and blogging (RMarkdown, blogdown)
- **AWESOME** community

...and why you should learn it

- Freeware, open - source & platform - independent
- **HUGE** choice of analytics and statistics libraries
- Integrates many data - sources
- Efficient data structures (data.frame, tibble, data.table, etc.)
- Produces arguably best Data Visualizations
- Can build interactive web apps (Shiny)
- **Easy** reporting and blogging (RMarkdown, blogdown)
- **AWESOME** community

...and why you should learn it

- Freeware, open - source & platform - independent
- **HUGE** choice of analytics and statistics libraries
- Integrates many data - sources
- Efficient data structures (data.frame, tibble, data.table, etc.)
- Produces arguably best Data Visualizations
- Can build interactive web apps (Shiny)
- **Easy** reporting and blogging (RMarkdown, blogdown)
- **AWESOME** community

...and why you should learn it

- Freeware, open - source & platform - independent
- **HUGE** choice of analytics and statistics libraries
- Integrates many data - sources
- Efficient data structures (data.frame, tibble, data.table, etc.)
- Produces arguably best Data Visualizations
- Can build interactive web apps (Shiny)
- **Easy** reporting and blogging (RMarkdown, blogdown)
- **AWESOME** community

...and why you should learn it

- Freeware, open - source & platform - independent
- **HUGE** choice of analytics and statistics libraries
- Integrates many data - sources
- Efficient data structures (data.frame, tibble, data.table, etc.)
- Produces arguably best Data Visualizations
- Can build interactive web apps (Shiny)
- **Easy** reporting and blogging (RMarkdown, blogdown)
- **AWESOME** community

...and why you should learn it

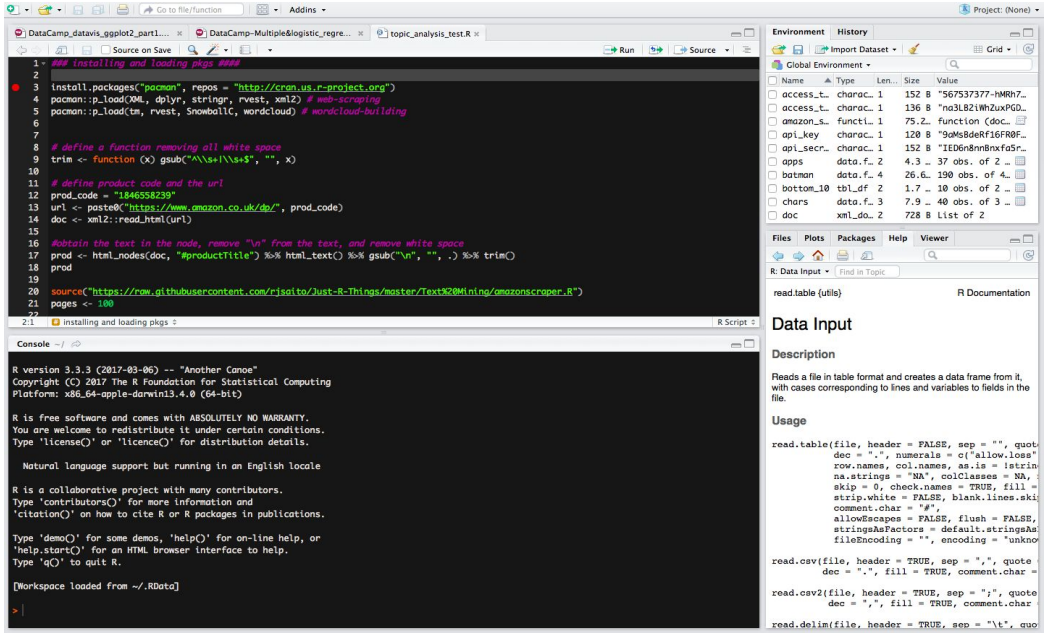
- Freeware, open - source & platform - independent
- **HUGE** choice of analytics and statistics libraries
- Integrates many data - sources
- Efficient data structures (data.frame, tibble, data.table, etc.)
- Produces arguably best Data Visualizations
- Can build interactive web apps (Shiny)
- **Easy** reporting and blogging (RMarkdown, blogdown)
- **AWESOME** community

Let’s learn some R!

- Installing R and RStudio
- RStudio Projects
- Data import and exploration
- Data manipulation & basic Data Visualisation
- Taking R to the next level

Installing R and RStudio

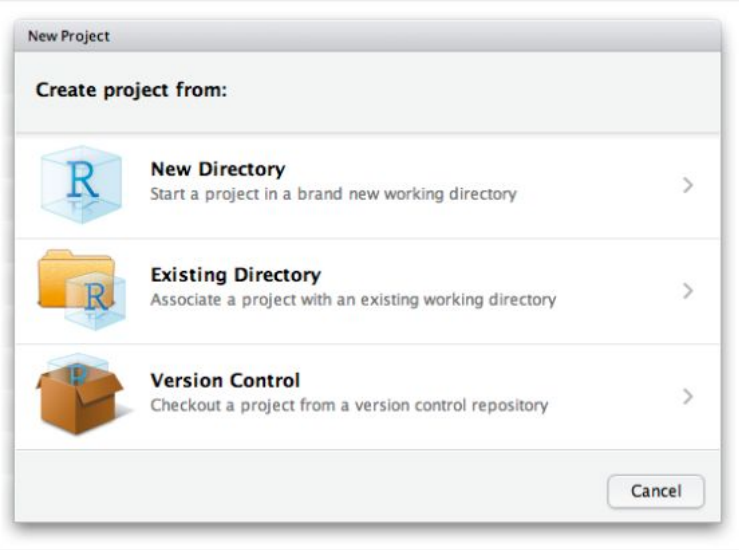
1. Install R from CRAN (https://cran.r-project.org/)
2. Install R's IDE: RStudio (https://www.rstudio.com/)
3. Open RStudio Console




YOUR
TURN!

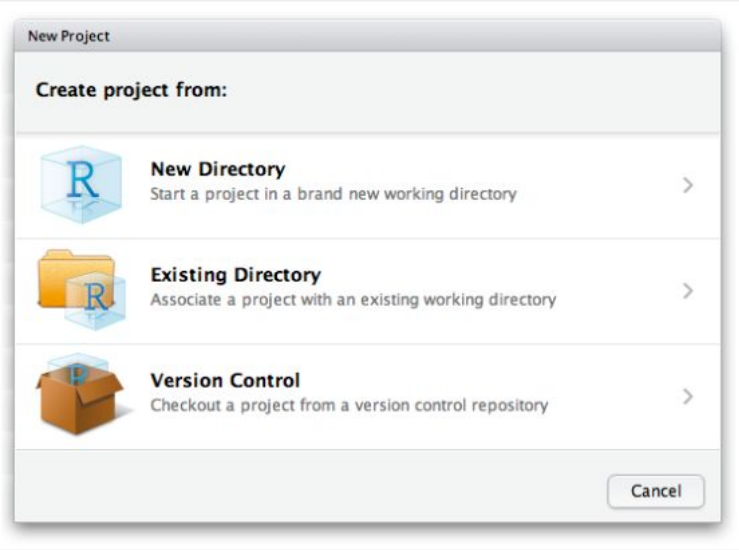
RStudio Projects

R Projects - divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents



RStudio Projects

R Projects - divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents

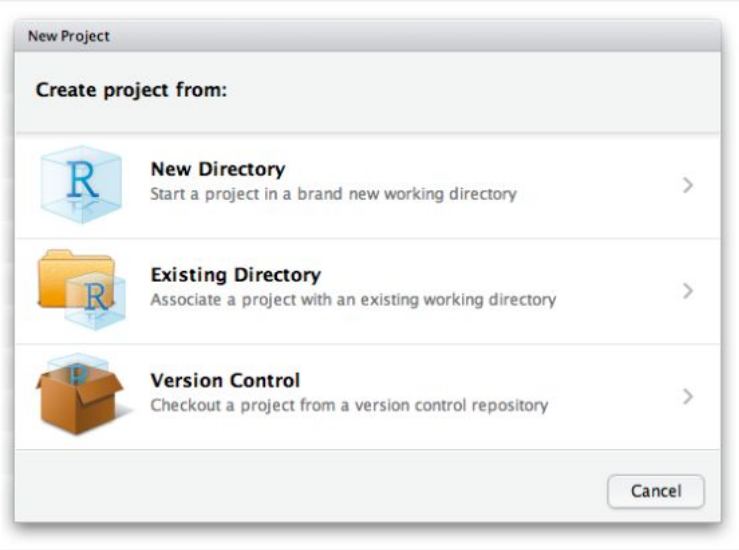



```
# shows current working directory
getwd()
```

```
# changes current working directory
setwd()
```

RStudio Projects

R Projects - divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents

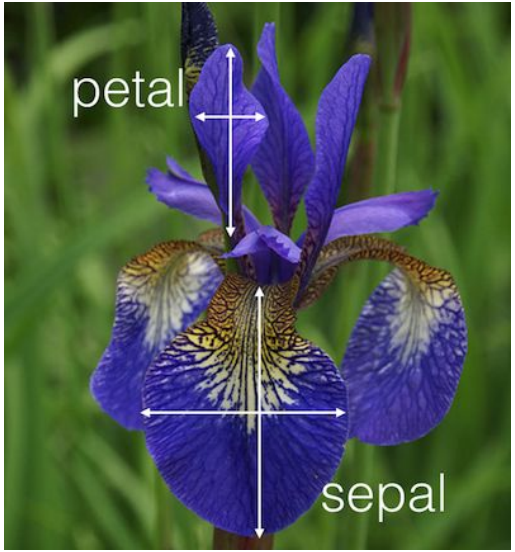



```
# shows current working directory
getwd()

# changes current working directory
setwd()
```


YOUR
TURN!

Data import and exploration

```
## In - built datasets

data(iris) # loads data into the workspace

head(iris) # views the top 6 rows

head(iris, 10) # views the top 10 rows

tail(iris, 10) # views bottom 10 rows

str(iris) # explores data structure

summary(iris) # summarises data
```




YOUR
TURN!

Data import and exploration

```
## saving and importing .csv files

# saves a data.frame in a .csv file
write.csv(iris, file = “20171019_iris_local.csv”)

# imports a .csv file and saves it in a new data object
local_iris <- read.csv(“20171019_iris_local.csv”)
str(local_iris)
```


YOUR
TURN!

Data manipulation & basic Data Visualisation

```
# numeric to integer

local_iris$Petal.Length
<-as.integer(local_iris$Petal.Length)


# numeric to character

local_iris$Petal.Width
<-as.character(local_iris$Petal.Width)
```

Data manipulation & basic Data Visualisation

```
# new logical var

local_iris$is_setosa <- local_iris$Species == "setosa"

# new numeric variables

local_iris$sepal_sum <- local_iris$Sepal.Width +
local_iris$Sepal.Length

local_iris$petal_sum <- local_iris$Petal.Width +
local_iris$Petal.Length
```


YOUR
TURN!

Data manipulation & basic Data Visualisation

```
## simple scatter plots in base R

plot(iris$Sepal.Length ~ iris$Sepal.Width) # basic plot

# base plot + main title

plot(iris$Sepal.Length ~ iris$Sepal.Width, main = "Sepal length by
sepal width")

# base plot + main title + axis labels

plot(iris$Sepal.Length ~ iris$Sepal.Width, main = "Sepal length by
sepal width", ylab = "Sepal length", xlab = "Sepal Width")
```


YOUR
TURN!

Data manipulation & basic Data Visualisation

```
## simple bar plots in base R

boxplot(iris$Sepal.Length ~ iris$Species) # base plot

# base plot + main title

boxplot(iris$Sepal.Length ~ iris$Species, main = "Sepal length by
species")

# base plot + main title + axis labels

boxplot(iris$Sepal.Length ~ iris$Species, main = "Sepal length by
species", ylab = "Sepal length", xlab = "Species")
```


YOUR
TURN!

Need help?

Each function comes with a documentation page that can be visualised typing:

- `help(name_of_function)` or
- `?name_of_function`

Other suggestions:

- Join R-Ladies - already done!
- Google your problem
- Browse http://rseek.org/ to find out which packages are available for a given topic
- Join an R users forum (e.g. R-help-archive - Google Groups or the RStudio community forum https://community.rstudio.com/)
- Post a question on https://stackoverflow.com

Taking R to the next level

Cheat Sheets & Reference Guides

- R Reference Card (http://cran.r-project.org/doc/contrib/Short-refcard.pdf)
- Writing R extensions (http://bit.ly/1H0U02a)
- Google's R Style Guide (https://google.github.io/styleguide/Rguide.xml)
- RStudio cheatsheets:
  - Data Visualization (http://bit.ly/1Foy1Lb)
  - Package Development (http://bit.ly/1CfbTD2)
  - Data Wrangling (http://bit.ly/1y2nh3f)
  - R Markdown (http://bit.ly/1BluuT5)
  - R Markdown Reference Guide (http://bit.ly/1L2tC7U)
  - Shiny (http://bit.ly/1GiGArG)

Taking R to the next level

Great tutorials:

- edx MiT course (https://www.edx.org/course/analytics-edge-mitx-15-071x-2)
- DataCamp (https://www.datacamp.com/)
- Coursera (https://www.coursera.org/learn/r-programming)
- Kaggle Tutorial (https://www.kaggle.com/mrisdal/titanic/exploring-survival-on-the-titanic)