

# Intro to R and R Markdown

DSC 152 Discussion – Week 1

# Agenda

- Quick discussion of R and RStudio's setup, R Markdown files
- Quick install of R and RStudio
- Live demos/coding in R
- Demo knitting code in R Markdown

# R and RStudio

- R – programming language, where you do computations
- RStudio – IDE (interface), where you can write/run R Code

# Installation Instructions for R

- Go to [www.r-project.org](http://www.r-project.org). Under “Getting Started”, click on "download R".
- Select a location from which you will download R (these are called CRAN mirrors); I would choose a site within the US. If one location is having trouble, choose another.
- Click on the download appropriate for your operating system (Linux, Mac, Windows).
- If you chose Windows in step 3), then you’ll want to click on “install R for the first time”, and then “Download R 4.6.0 for Windows” (or whatever the most recent version number is when you are reading these instructions). If you chose Mac in step 3), then you’ll click on R-4.6.0.pkg near the bottom of the screen. At the end of this step, you should be downloading a file containing the R language.
- Once the file is downloaded, you can double-click it and begin the installation process. When you’re done installing R, then you need to download and install RStudio.

# Installation Instructions for RStudio

- Go <https://posit.co/download/rstudio-desktop/> and go to the next step.
- Scroll down a little to the “All Installers and Tarballs” and find the Installer file appropriate for your operating system. At the end of this step, you should be downloading RStudio.
- Once the file is downloaded, you can double-click it and begin the installation process. When you’re done installing RStudio, you should be able to run RStudio and type `print(“Hello World”)` in the console.
- After hitting enter, you should get `[1] “Hello World”`

# Typical Setup in RStudio

The image shows the RStudio interface with four key components highlighted by colored boxes and arrows:

- Script editor:** A blue box highlights the top-left pane containing R code:

```
1 print_num <- function(n) {  
2   print(n)  
3 }
```
- Environment:** A red box highlights the top-right pane showing the Environment tab. It displays the Global Environment with a data object named 'normalSamples' (Large matrix (230000 elements, 1.8 MB)). The 'Values' section lists variables: best\_season (59), goals (num [1:21] 12 20 28 49 12 4 17 30 31 25 ...), meanEstimates (num [1:10000] 110.7 104.7 108.4 103.9 99.3 ...), mu (106), n (23), N (10000), sample (num [1:23] 113.7 88.6 105.7 112.2 102.1 ...), and sigma (11.916375287813).
- Console (terminal):** A green box highlights the bottom-left pane showing the R console output, including the R version (4.5.1), copyright information, and a list of help commands like 'license()', 'demo()', and 'q()'. The prompt '>' is visible at the bottom.
- Files/Plots/Packages:** A purple box highlights the bottom-right pane, which is currently empty, representing the area for file explorer, plots, and package management.

# Directories and Packages

## Working Directory:

- Folder R 'looks in' by default
- Check it with `getwd()`
- Set/change it with `setwd('path/to/folder')`
- Set/change it in RStudio's menu (Session → Set Working Directory and then choose your directory)

## Packages (like those in Python):

- Bundles of extra functions written by community
- Install each package once:  
`install.packages("ggplot2")`
- Load necessary packages in each session:  
`library(ggplot2)`
- Packages lives on CRAN ('app store' for R)

# R Markdown

- `.Rmd` file combines text, code, LaTeX, and outputs into one file
- **Knit** is how RStudio runs code and compiles everything into a file (pdf, docx, HTML, etc.)
- We ask you submit your HWs via R Markdown files as pdf outputs

```
11 ## R Markdown
12
13 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word
14 documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
15 When you click the Knit button a document will be generated that includes both content as well as the
16 output of any embedded R code chunks within the document. You can embed an R code chunk like this:
17 ```{r cars}
18 summary(cars)
19 ```
20
```

Raw .Rmd file content

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
## Min.   : 4.0    Min.   :  2.00
## 1st Qu.:12.0    1st Qu.: 26.00
## Median :15.0    Median : 36.00
## Mean   :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
## Max.   :25.0    Max.   :120.00
```

Produced output in PDF

# R Markdowns & LaTeX

- Knitting R Markdown files as pdf files requires local LaTeX compiler
- Local LaTeX installation is difficult because of size
- Easiest solution is `tinytex` (more lightweight LaTeX distribution designed for R)
- Install once with two commands:
  - `install.packages("tinytex")`
  - `tinytex::install_tinytex()`
- After this, knitting works as intended

# Demos in R

# Couple Problems in R

1. Suppose you collected temperatures in San Diego across six days, recording 72, 79, 87, 68, 73, and 73 degrees Fahrenheit. Store these observations in a vector `temps`. Use `temps` to find the mean and maximum temperature, and then write an expression to check that your vector has exactly 6 observations.
2. Using `temps`, create a new vector `hot_days`, that contains only temperatures that were at least 75° Fahrenheit. How many such days were there? What was the temperature on those days?
3. Using anonymous data for past student performance in a particular DSC course, researchers believe a student's grade is a normally distributed random variable with mean 83.8 and variance 27.2. If the standard cut-off for an A- is 90.0, what is the approximate proportion of students that received an A- or better? (Hint: you'll need to use one of the `_norm` functions we saw.)
4. Load in the built-in `cars` dataset. First, using the function `nrow`, determine how many rows are in the dataset. Then, create a vector `speed_kph` which is the speed in kilometers per hour for each observation in the dataset (the conversion rate is 1mph = 1.609kph). What are the first few values of `speed_kph`, and how many speeds exceed 40kph?

# Demo in R Markdown