

IntroR

Prof. Eric A. Suess

January 27, 2020

R Markdown

This is an R Markdown presentation. 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>.

File > New File > R Markdown..., select Presentation and then HTML(ioslides), be sure to add the Title: and Author:

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.

Today

- ▶ We will introduce R
- ▶ Data structures in R
- ▶ Functions in R
- ▶ Understanding data with R
- ▶ R Packages and Libraries
- ▶ R Scripts, R Notebooks, R Projects

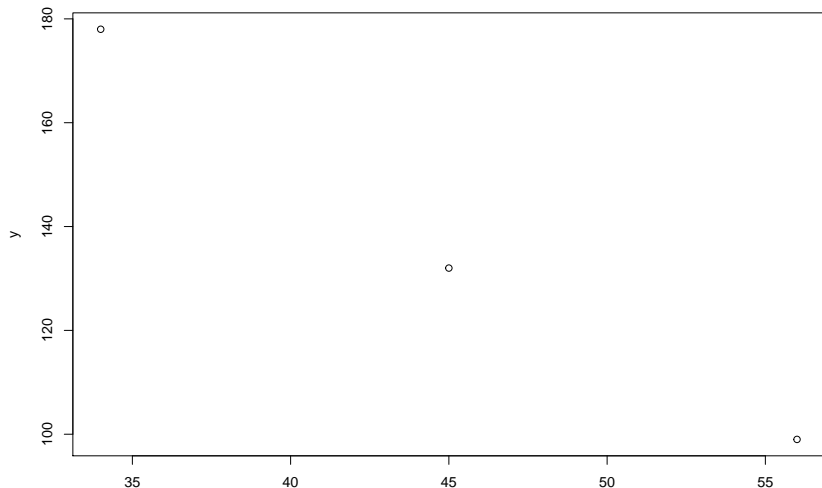
Introduction to R: Data Structures

The main types of data structures in R

- ▶ *vectors* - numeric or character or logical
- ▶ *factors* - for nominal variables/features
- ▶ *lists* - numeric and/or character and/or logical
- ▶ *data frames* - list of vectors and/or lists
- ▶ *matrices* - numeric, r by c , fills columns
- ▶ *arrays* - layers, like sheets in MS Excel

Introduction to R: vectors

```
x <- c(34,45,56)
y <- c(178,132,99)
plot(x,y)
```



Introduction to R: factors

```
gender <- factor(c("F", "M", "F"))  
gender
```

```
## [1] F M F
```

```
## Levels: F M
```

Introduction to R: lists

```
subject1 <- list(x = x[1], y = y[1],  
  gender = gender[1])  
subject1
```

```
## $x  
## [1] 34  
##  
## $y  
## [1] 178  
##  
## $gender  
## [1] F  
## Levels: F M
```

Introduction to R: data frames

```
mydata <- data.frame(x, y, gender)
mydata
```

```
##      x    y gender
## 1 34 178      F
## 2 45 132      M
## 3 56  99      F
```

```
mydata$x
```

```
## [1] 34 45 56
```

```
mydata$gender
```

```
## [1] F M F
## Levels: F M
```


Introduction to R: data frames

```
mydata <- data.frame(x, y, gender)
mydata[1,]
```

```
##      x    y gender
## 1  34 178      F
```

```
mydata[,c(2,3)]
```

```
##      y gender
## 1 178      F
## 2 132      M
## 3  99      F
```

Introduction to R: matrices

```
X <- matrix(c(x,y), ncol=2)  
X
```

```
##      [,1] [,2]  
## [1,]   34  178  
## [2,]   45  132  
## [3,]   56   99
```

Introduction to R: Managing data

Set the working directory.

```
> getwd()
```

```
> setwd("C:\\ path to where your data is, with double \\")
```

In RStudio try to set the working directory in one of three ways.

1. Session > Set Working Directory > Choose Directory
2. Files browse and More > Set As Working Directory
3. Best practice is to use an **R Project** and the *here* package.

Introduction to R: Managing data

Reading and writing .csv files

```
> usedcars <- read.csv("usedcars.csv", stringsAsFactors = F)
```

```
> write.csv("mydata", file "mydata.csv")
```

In RStudio try to load the data with the

Environment > Import Dataset >

From Text (base)... or From Text (readr)...

Introduction to R: Understanding data

When exploring **quantitative/numeric variables** we use

- ▶ mean and median
- ▶ standard deviation
- ▶ 5-number summary
- ▶ box-plots
- ▶ histograms
- ▶ normal distributions?

Introduction to R: Understanding data

```
usedcars <- read.csv("usedcars.csv",  
  stringsAsFactors = FALSE)  
head(usedcars)
```

```
##   year model price mileage  color transmission  
## 1 2011   SEL 21992   7413 Yellow          AUTO  
## 2 2011   SEL 20995  10926  Gray          AUTO  
## 3 2011   SEL 19995   7351 Silver         AUTO  
## 4 2011   SEL 17809  11613  Gray          AUTO  
## 5 2012    SE 17500   8367  White         AUTO  
## 6 2010   SEL 17495  25125 Silver         AUTO
```

Introduction to R: Understanding data

```
usedcars <- read.csv("usedcars.csv",  
  stringsAsFactors = FALSE)  
summary(usedcars$price)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      3800   10995   13592   12962   14904   21992
```

Introduction to R: Understanding data

```
mean(usedcars$price)
```

```
## [1] 12961.93
```

```
sd(usedcars$price)
```

```
## [1] 3122.482
```

```
range(usedcars$price)
```

```
## [1] 3800 21992
```


Introduction to R: Understanding data

When exploring **qualitative/categorical variables** we use

- ▶ counts and percentages
- ▶ tables
- ▶ mode
- ▶ bar graphs

Introduction to R: Understanding data

When exploring the **relationships** between **quantitative/numeric variables** we use

- ▶ correlation
- ▶ scatterplots

Introduction to R: Understanding data

When exploring the **relationships** between **qualitative/categorical variables** we use

- ▶ tables
- ▶ Chi-Square

Introduction to R: Libraries

We will be using the *gmodels* library during the class. Install the package.

```
> install.packages("gmodels")
```

```
> library(gmodels)
```

In RStudio

Packages > Install

Introduction to R: Scripts

The author of our book used R Scripts.

File > New File > R Script

R Scripts has the file extension .R

Introduction to R: Notebooks

RStudio offers R Notebooks. With R Notebooks you can blend the use of R code with your own text.

File > New File > R Notebook

To use the R Notebook you can add code chunks with

Ctrl + Alt + i

R Notebooks has the file extension .Rmd

The md stands for markdown.

Introduction to R: Projects

As a best practice it is recommended that you create an **R Project** for each new R program you work on.

File > New Project

Or to the right click on

Project (None) > New Project

In the directory where you create your project there is a file with the extension `.Rproj`

Using R Projects keeps all your related files together and make the reading and writing of files easier.

Introduction to R: DataCamp Code School

- ▶ DataCamp
- ▶ Introduction to R