R for Beginners

Barbara Fulda

12th December 2014

What lies ahead?

- The why question
- The what question
- The how question

Why R?

- Open-Source: It's free! (http://www.r-project.org)
- Cutting-edge analytics: Excellent domain specific support a lot of people from different countries work with R and constantly publish new content
- A programming language designed by and for statisticians

Why not stay with Excel?

- R code makes it easy to document and reproduce your analysis
- Therefore sharing and working together on a project is much easier
- Any form of data can be used (.csv, data you get through scraping websites etc.)
- R can do more than Excel as there exist about 5000 packages and ever more are developed continuously
- Data visualization: Create easily customizable and beautiful graphs

What are the essential ingredients of R?

- Input: Write code in the script window
- Output: The results are in the output window
- Graphs are displayed in another window

We will quickly try this out. . .

- Open the R script labelled "beginners.R"
- Select the lines of code below the line "R as a calculator"
- Press "Run"

What is an R editor?

- ► Here you use R studio
  -> Easy and nice user surface
- ► There are many other options
- ► Tinn-R
- ► Vim
- ► Emacs etc.

What does this mean for analysis?

- Numbers and characters

```
1+1
```

```
## [1] 2
```

```
r "Hello, World!"
## [1] "Hello, World!"
```

Wait! Before we begin, we store the commands in a file

- Where are we?

r getwd()

- Where do we want to be?

r setwd("/home/jim/psych/risk/adol")

- As with any data anlysis software we can store our commands in a file called "R-script"

Comment it for later use-> #Hello

Basics (1)

- Objects (variables etc.)

```
x<-2
```

- What is x again?

```
x
```

```
## [1] 2
```

- "Ah, typed it wrong..."

```
x<-5
```

Basics (2)

```
y<-4
```

```
z=x+y
z
```

```
## [1] 9
```

Load packages for specialized use

- Some packages are preinstalled
- any other package is installed by you

```r
install.packages("ggplot2")
```

- Why?
  It is sparse. Storage remains empty and calculations are quick

When I am stuck. . . What to do?

```
help(sum)
```

```
example(min)
```

Load data for analysis

- ▶ We use the example data sets available within R

```
library(MASS)
data()
```

- ▶ We will use the "ToothGrowth" dataset

```
data(ToothGrowth)
```

- ▶ You can load several datasets at the same time and use them simultaneously

```
data(Animals)
```

- ▶ Look at the environment!

Simple calculations with R

```
x<-mean(ToothGrowth$len)
x
```

```
## [1] 18.81333
```

```
y<-mean(ToothGrowth$dose)
y
```

```
## [1] 1.166667
```

```
x/y
```

```
## [1] 16.12571
```

How to access your data

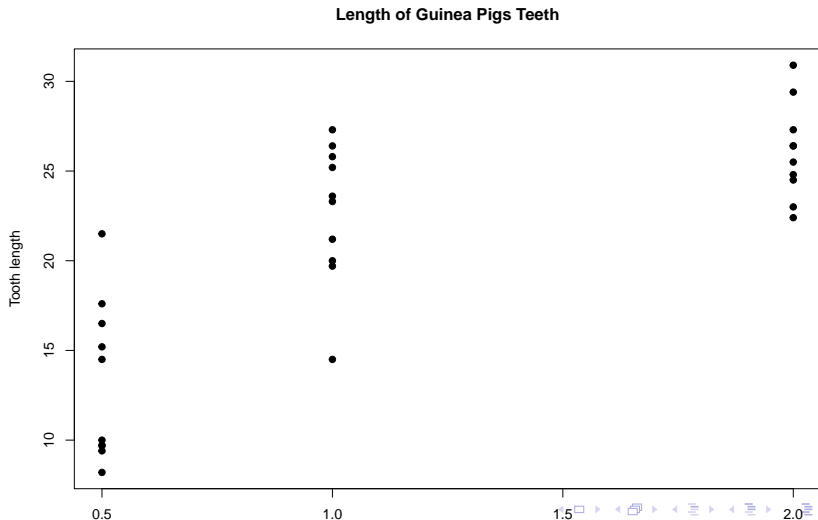- What's the content of a variable?

```
ToothGrowth$len
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2 11.2  5.2  7.0 1
## [15] 22.5 17.3 13.6 14.5 18.8 15.5 23.6 18.5 33.9 25.5 2
## [29] 23.3 29.5 15.2 21.5 17.6  9.7 14.5 10.0  8.2  9.4 1
## [43] 23.6 26.4 20.0 25.2 25.8 21.2 14.5 27.3 25.5 26.4 2
## [57] 26.4 27.3 29.4 23.0
```

Plotting data (1)

- ▶ The Effect of Vitamin C on Tooth Growth in Guinea Pigs
- ▶ They were given Orange Juice or ascorbic acid.
- ▶ How long do their teeth grow when they are given orange juice?
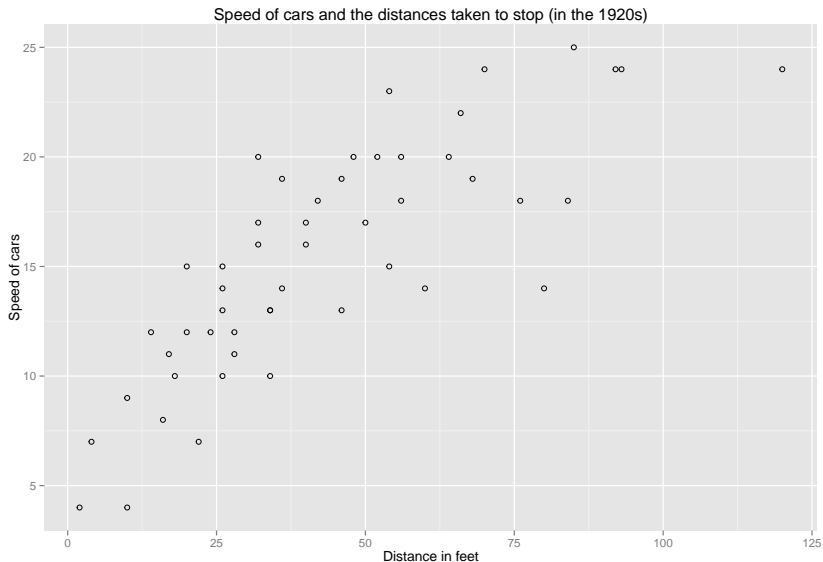
Length of Guinea Pigs Teeth

Plotting data (2)

Plotting a regression fitline
- Cars and the distance they travel depending on speed
- First view the data

```
library(ggplot2)
data(cars)
head(cars, n=5)
```

```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
```
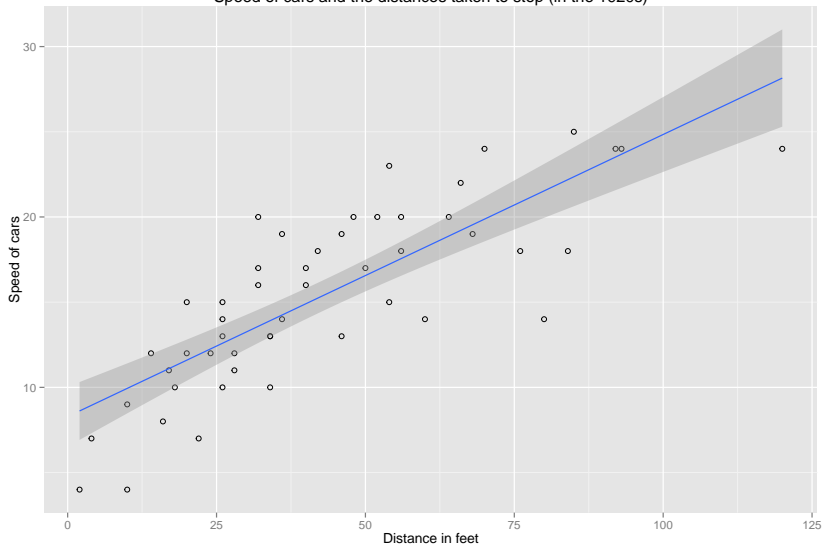
Plotting data (3)

Speed of cars and the distances taken to stop (in the 1920s)
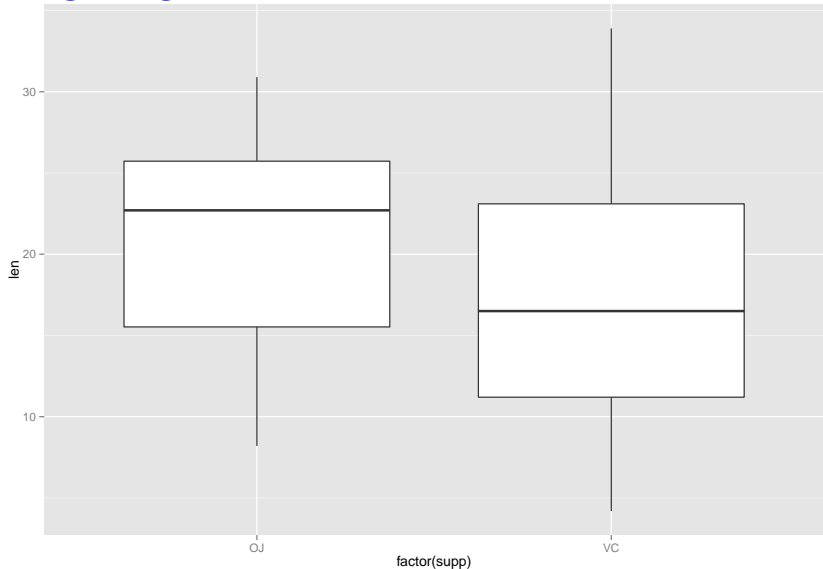
Plotting Data (4)

-Regression fit line

Speed of cars and the distances taken to stop (in the 1920s)

Plotting categorical data

Orange Juice clearly lets guinea pigs teeth grow
longer

Thank you!

Questions?

Where to continue

Here are some websites to continue learning R

Introductory courses:
https://www.codeschool.com/courses/try-r
https://www.coursera.org/course/compdata
https://www.coursera.org/course/datasci
http://sentimentmining.net/StatisticsWithR/