Datapolitan

Data Solutions for the Modern Metropolis

tiny PANTHER

Data Analysis with R

Follow along at: http://bit.ly/data-analysis-r

See the code at: http://bit.ly/data-analysis-r-code

Open your browser:

Go to this link: **http://student___.datapolitan.com**

Username: **rstudio** Password: **rstudio**

Julia Marden

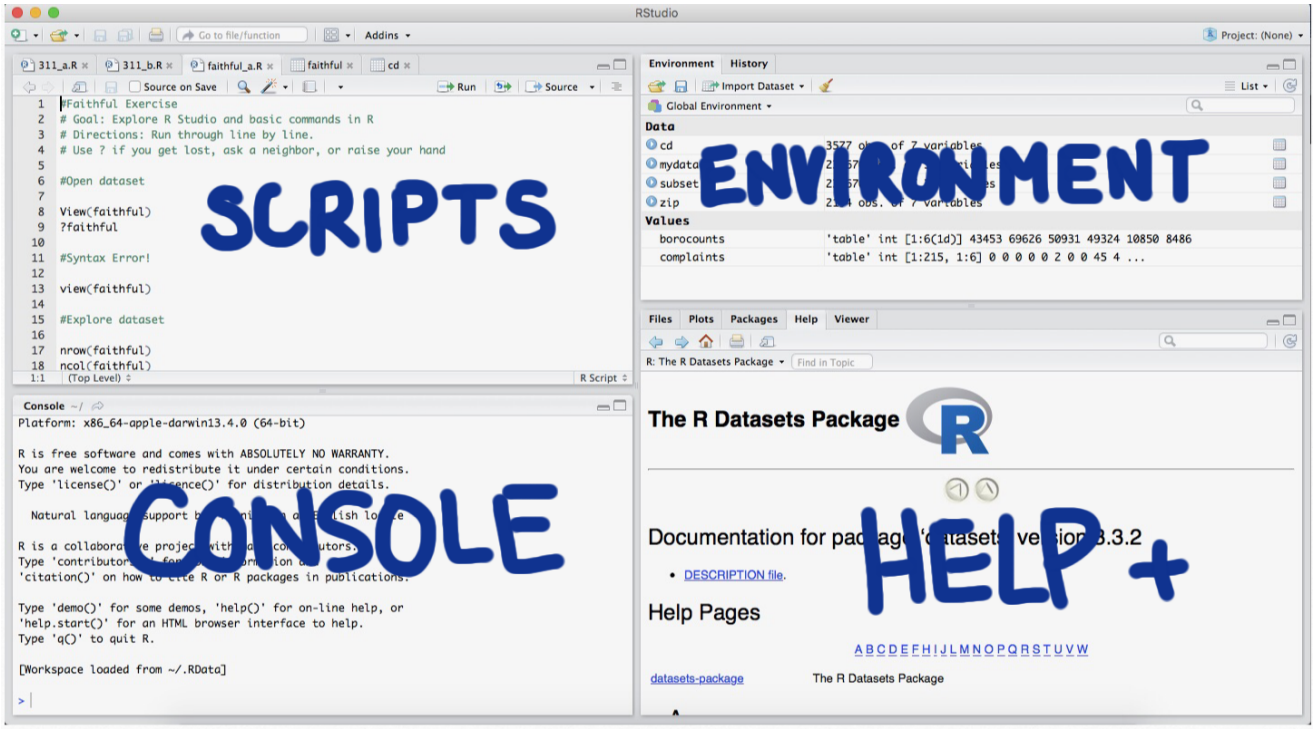
- Email: julia@tinypanther.com
- Website: http://tinypanther.com
- Twitter: @juliaem

Richard Dunks

- Email: richard@datapolitan.com
- Blog: http://blog.datapolitan.com
- Twitter: @datapolitan

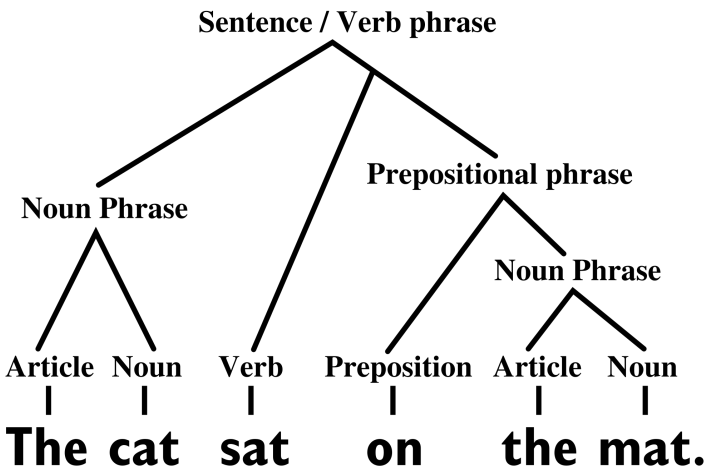
Resources

- Stat Methods (http://statmethods.net) - Great documentation for doing data analysis in R
- UCLA Stats (http://www.ats.ucla.edu/stat/) - Many examples of statistical analysis with comparisons between R, Stata, SPSS, etc.
- Swirl (http://swirlstats.com) - install.packages("swirl")
- R Cookbook (http://www.cookbook-r.com/) - Free online walkthrough of the basics



What is Syntax?

Basic constituent structure analysis of a sentence:



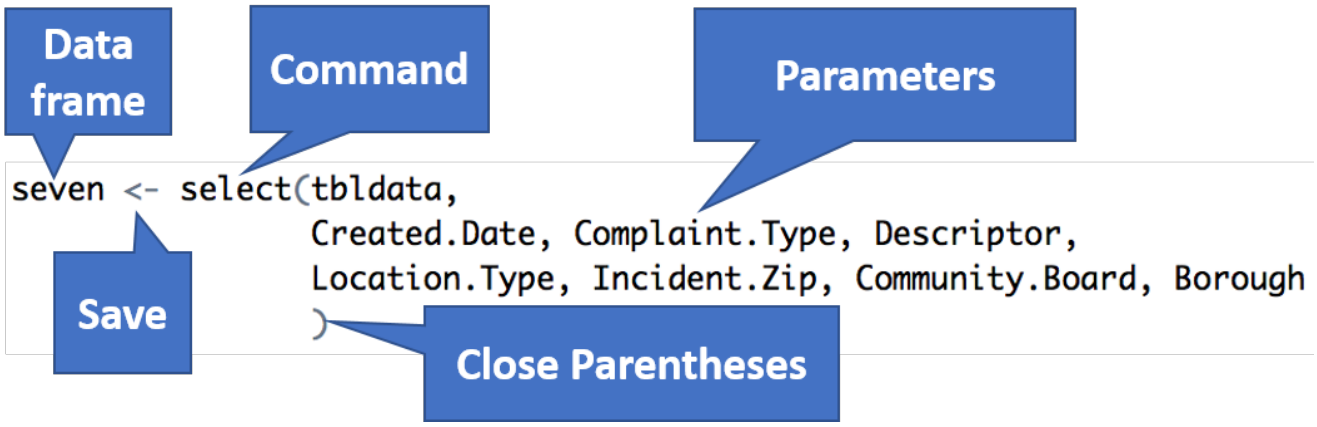
R Syntax

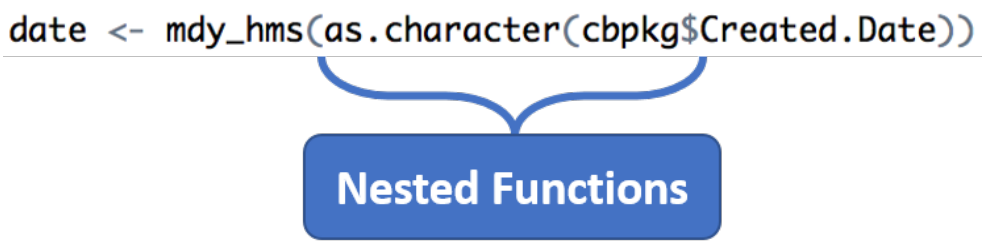
```
# basic command

command(dataset)
View(faithful)
```

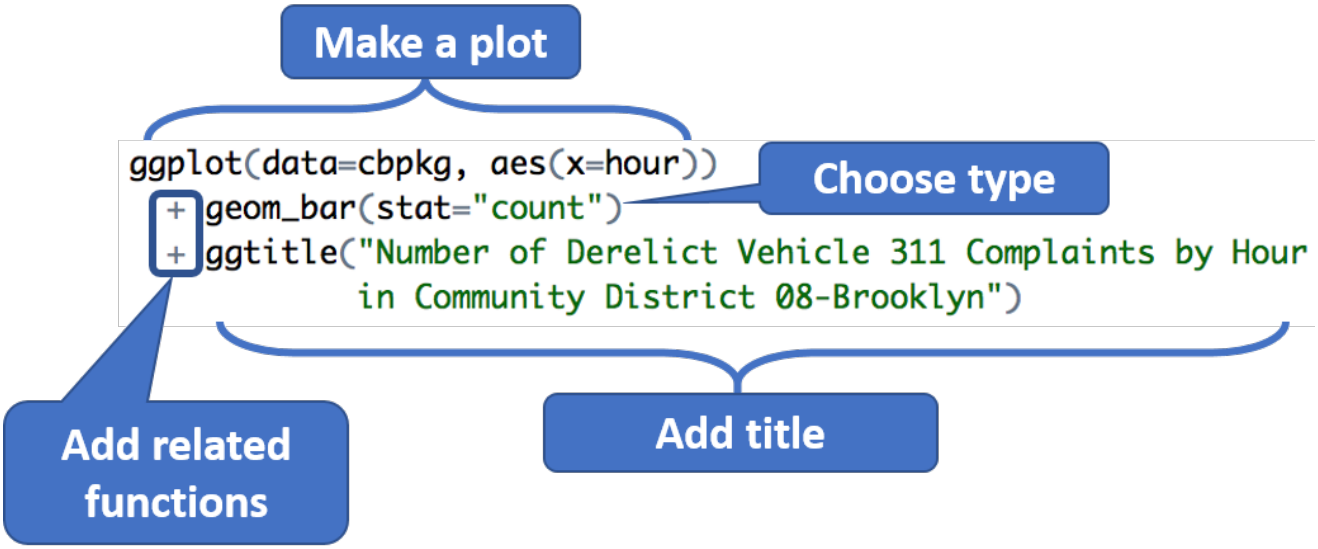
```
# select a column

command(dataset$column)
mean(faithful$waiting)
```

```
# get help

?help
?faithful
```

Important Parts of R Commands We're Using







Exploring Data

```
View()
# show dataset as spreadsheet in Viewer
```

```
str()
# identify data type and structure
```

```
nrow()
# identify the number of rows
```

```
ncol()
# identify the number of columns
```

```
colnames()
# list the name of every column
```

Visualizing Data

```
hist()
# make a chart with numeric data
```

```
plot()
# plot two numeric variables along an x-y axis
```

```
abline()
# add a trendline to a plot
```

```
table()
# make a table with factor data
```

```
prop.table()
# make a table with percentages
```

```
barplot()
# make a chart with factor data
```

Manipulate Data

```
sort()
# sort the values in a column
```

```
data.frame()
# structure data into a matrix
```

```
subset()
# extract data from a dataframe
```

Calculating Summary statistics

```
min()
# identify minimum value
```

```
max()
# identify maximum value
```

```
median()
# calculate median value
```

```
mean()
# calculate mean value
```

Links

- Download R: https://cloud.r-project.org/
- Download RStudio: https://www.rstudio.com/products/rstudio/download/
- Download exercise files from this class: http://bit.ly/data-analysis-r-code

dplyr

```
install.packages("dplyr")
require(dplyr)
```

```
tbl_df()
#create a dataframe
```

```
filter()
select()
# create a subset; filter for rows, select for columns
```

```
mutate()
# add a column
```

```
arrange()
# sort rows by category
```

ggplot2

```
ggplot()
#plot a dataframe
```

```
geom_bar()
# make a proportional bar chart
# alternative is geom_col()
# used for factor data
```

```
ggtitle()
# add a title to a plot
```

lubridate

```
install.packages("lubridate")
require(lubridate)
```

```
mdy_hms()
#format timestamp into month, day, year, hour, min and second
# other commands: mdy_hm, mdy, dmy, etc.
```

```
hour()
# extract hour from timestamp
# other commands: day, minute, second, etc.
```