Datapolitan

Data Solutions for the Modern Metropolis

# Overview of Data Analysis with R

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

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

## What to Expect Today

- 9:15 – Welcome and Introductions
- 9:25 – What is R?
- 10:40 – 10 min break
- 10:50 - Exercise
- 12:30 – Dismissal

## What Not To Expect Today

- Becoming a R expert
- Becoming a data analytics pro
- Becoming a visualization wizard

## What is R?

- Statistical programming language
- Open-source
- Made for and by people who work with data

## R vs. Excel

- Language vs. program
- Big data
- Different structures and data types
- More potential

## R Data

- Vectors and matrices
- Lists and data frames
- Numeric vs. factor

Your student number is:_____

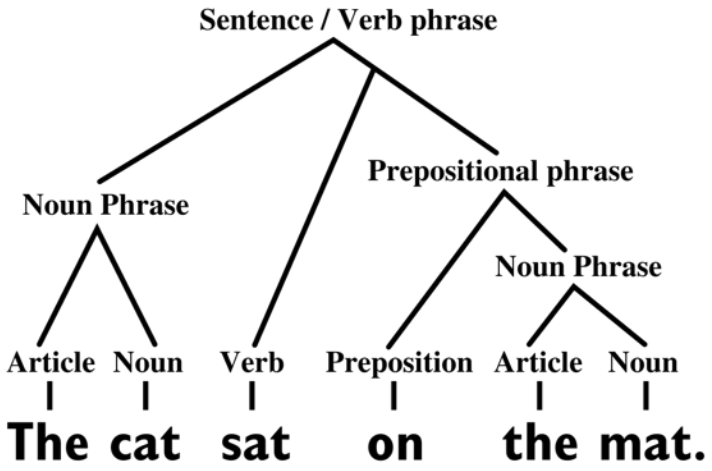
Your weblink is: **http://student__.datapolitan.com**

Username: **rstudio**

Password: **rstudio**

## 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
```

## 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
```

## What we've covered

- R as a language
- RStudio
- Open a dataset
- Explore a dataset
- Visualize a dataset

## Final Thoughts

- R is a powerful tool for cleaning, analyzing, and visualizing data
- Integrating it into your workflow takes practice and a commitment to not giving up (Google is your friend)
- RStudio makes it easy to get started
- You should be able to download RStudio on your work computer (Use the zip/tarball option)


## 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.
- Stack Overflow (http://stackoverflow.com/)- One of the best Q&A sites for technical questions of all kinds
- R Cookbook (http://www.cookbook-r.com/) - Free online walkthrough of the basics