R for Data Analysis

TECH TALENT ACADEMY

TECH TALENT ACADEMY

Session Content

Data Visualisation

In-built Data Sets

Continuous v Discreet Data

WOMEN IN DATA ACADEMY

Why visualise data?

Data is only valuable when we can glean insight from it. Raw data is a necessary tool for this, but it is chaotic and its “meaning” is not easily accessible.

Good data visualisation helps the human brain to quickly understand these insights, meaning that the data can be harnessed in a way that provides benefit.

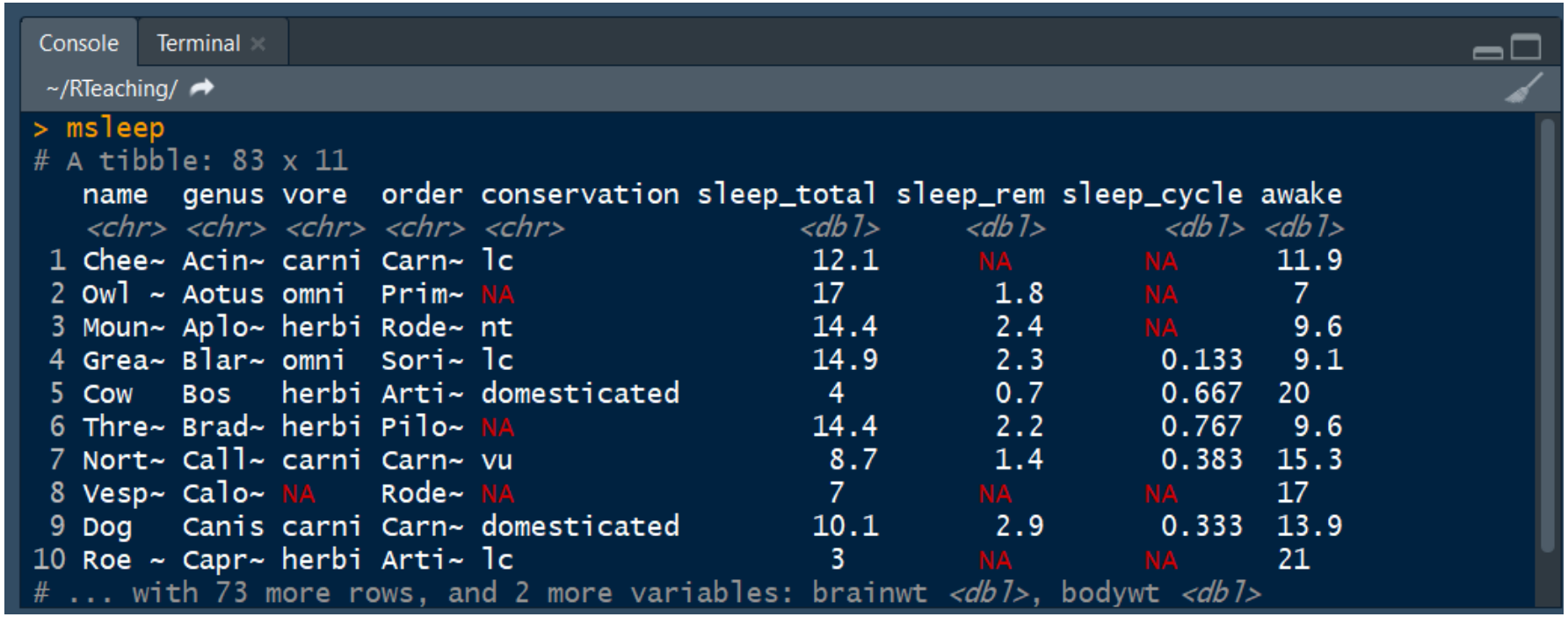
In-built Data Sets

RStudio has in-built sample data sets. These allow you to practice on samples of large data sets without having to import or build your own.

CODE	PURPOSE
data(package = .packages(all.available = TRUE))	All available datasets for all available packages
data()	All available datasets for currently installed packages
data(package = “package name")	Datasets available within specific package.

In-built Data Sets Continued

To view a small sample of the data for that data set, type the name of the data set into the console:



For more information about the data, type the name of the data set into the console preceded by a ?. For example:

?msleep

Ggplot data sets

diamonds → Prices of over 50,000 round cut diamonds

economics → US economic time series

economics_long → US economic time series

faithfuld → 2nd density estimate of Old faithful data

luv_colors → 'colors()' in Luv space. Mapping between assorted color spaces

midwest → Midwest demographics

mpg → Fuel economy data from 1999 to 2008 for 38 popular models of cars

msleep → mammals sleep dataset

presidential → Terms of 11 presidents from Eisenhower to Obama

seals → Vector field of seal movements

txhousing → Housing sales in Texas

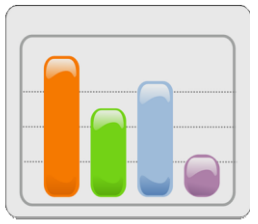
Graph Commands

The code to designate a graphic's style is the below:

geom = "bar"

The most frequently used geom options are:

"bar"	First tabulates frequencies of each value, then makes a barplot.
"histogram"	Makes a histogram.
"point"	Makes scatterplots.
"line"	Makes a line plot.
"boxplot"	Makes a boxplot.
"density"	Makes the density plot
"smooth"	Fits a smooth line to a cloud of points and plots the output.
"dotplot"	Makes a dotplot.



Continuous v Discreet Data

Discreet

Categorical data that can be classified firmly and distinctly.

Month of the year (Jan, Feb, March, April)

Test scores (A*, A, B, C)

Star rating (1, 2, 3, 4, 5)

Temperature (high, medium, low)

Continuous

Data that is not restricted to defined distinct values, can occupy any value within a continuous range.

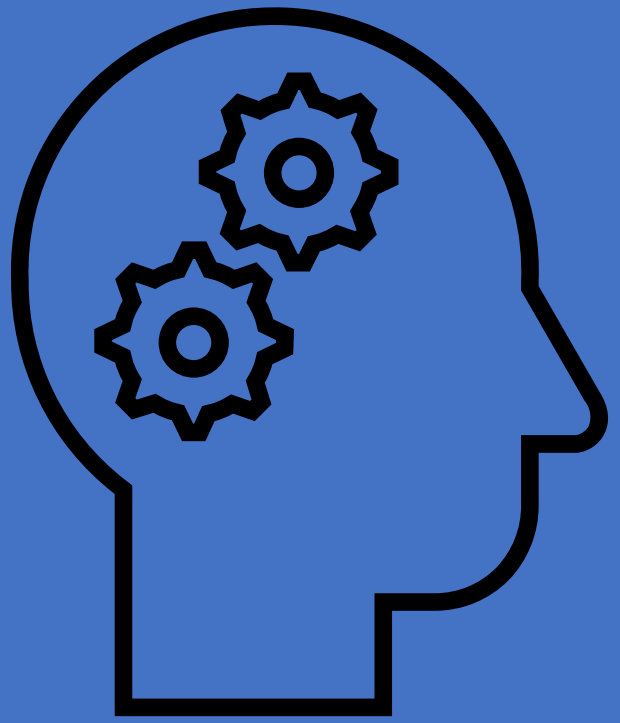
Date (10th July 2020, 11th July 2020, 12th July 2020)

Percentages (58%, 167%, 240%)

Length (1mm, 65mm, 1500mm)

Temperature (17.4°C, 23.7°C, 34.8°C)

Home Learning Project



Imagine the following scenario:

You are a data analyst/scientist at an organisation. You have been given a data set and asked to create a meaningful data visualisation using this data.

Using the ggplot in-built data sets in RStudio and the qplot function, get your creative juices flowing and create a meaningful and impactful data visualization using your preferred data set.

WOMEN IN DATA ACADEMY

WOMEN IN DATA ACADEMY

TECH TALENT ACADEMY