

R for Data Analysis | NTFAQ

Speakers: Melinda Ronca-Battist, Dorian Kvale, Kristie Ellickson

May 16, 2018

Follow along online

https://mpca-air.github.io/NTF_learn_R/NTF_Demo.html

Install R

https://mpca-air.github.io/NTF_learn_R/00_Install.html

Upcoming R Training!

- 2-Day hands-on course
- Using real air monitoring data
- Includes criteria pollutants and air toxics
- Make charts, maps, and pollution roses

Contact Melinda for more information.

Previous training materials :: R Camp!

Tutorials and workshops online at <https://MPCA-air.github.io/RCamp>

Resources for learning R

- RStudio: <https://www.rstudio.com/online-learning/>
- `library(swirl)` for interactive lessons in R
- Intro to R with Pirates! at <http://tryr.codeschool.com/>
- More pirates at YaRrr! Pirates Guide to R
- Coursera data science: <https://www.coursera.org/jhu>
- R for Data Science by Hadley - <http://r4ds.had.co.nz/>
- R Bloggers articles at <https://www.r-bloggers.com/>
- #rstats on Twitter
- R cookbooks at cookbook-r.com (a variety of simple recipes for data and charts)

A guide for Air data methods

We've started an online community guide for air data analysis methods. We use it as a resource to provide similar methodologies to calculate means, compare sites, and other common air data tasks. Everyone is invited and encouraged to contribute. It is shared online at <https://mpca-air.github.io/air-methods/>.

Shiny tools in presentation

Wind pollution roses:

<https://air-data.shinyapps.io/pollution-roses/>

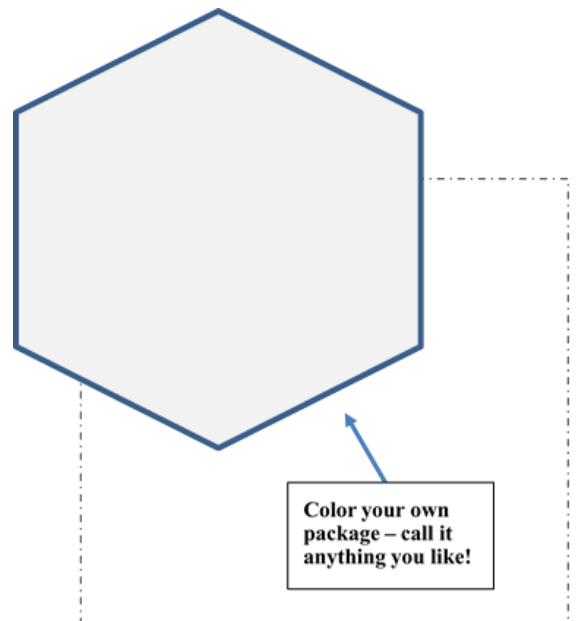
Key terms

package An add-on for R that contains new functions someone created to help you. It's like an App for your phone.

library The name of the folder that stores your R packages.

script A text file used to record the step-by-step instructions of your data analysis.

RStudio A helpful user interface that organizes your data, charts, R scripts and packages into one bundle.



R Packages in presentation

Core packages



• **ggplot2**, which implements the grammar of graphics. You can use it to visualize your data.



• **dplyr** is a grammar of data manipulation. You can use it to solve the most common data manipulation challenges.



• **tidyr** helps you to create tidy data or data where each variable is in a column, each observation is a row and each value is a cell.



• **readr** is a fast and friendly way to read rectangular data.

Data reading

- `library(readxl)`

Air specific analysis and charts

- `library(openair)`

Data processing and cleaning

- `library(tidyverse, lubridate)`

Charts

- `library(ggbeeswarm, waffle, ggpomological)`

Maps

- `library(leaflet, sf)`

Interactive tools

- `library(shiny)`

Just for fun

- `library(weatherAlerts, cowsay)`

Dynamic documents with charts

- `library(rmarkdown)`

Sample air data - data_2015

site_catid	Year	Date	Hour	Parameter	Conc	Latitude	Longitude
27-017-7417	2015	2015-06-10	22	88101	7	46.71	-92.51
27-017-7417	2015	2015-06-10	23	88101	9	46.71	-92.51
27-017-7417	2015	2015-06-11	0	88101	7	46.71	-92.51

Sample R commands

```
# Load Excel data
read_excel("air_data.xlsx")

# Scatterplot of Ozone vs. Temperature
ggplot(data, aes(y = O3_ppm, x = Temp_F)) + geom_point()

# Filter data
filter(data, Concentration > 99, site == "Big Woods Monitor")

# Add new column to data
mutate(data, O3_ppb = O3_ppm * 1000)
```

Contacts

Melinda Ronca-Battista: Melinda.Roncabattista@gmail.com

Kristie Ellickson: Kristie.Ellickson@state.mn.us

Dorian Kvale: Dorian.Kvale@state.mn.us