R Stats & Swirl

March 19 & 20, 2025

www.dartgo.org/RRADworkshops

Research Data Services
researchdatahelp@groups.dartmouth.edu
Stephen P. Gaughan
Research Facilitator - Geospatial Programmer/Analyst
603-646-9524
stephen.p.gaughan@dartmouth.edu
Hanover, NH 03755

DARTMOUTH

Data Frames, CSV files & Tibbles

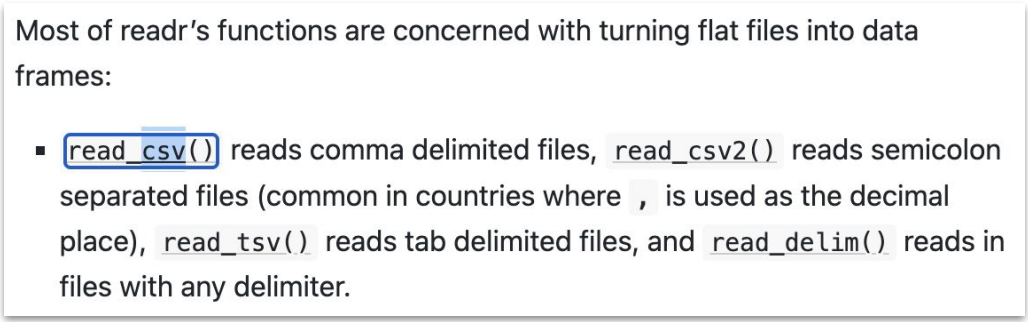
Data Frames, CSV's and Tibbles

Data frames are structured data objects in programming languages like R, used to store tabular data in rows and columns, where each column can contain different types of data (e.g., numeric, character).

Tibbles are a modern take on data frames in R, part of the tidyverse, designed to be more user-friendly and efficient, providing better printing and subsetting capabilities.

CSV files (Comma-Separated Values) are a widely-used plain text file format for storing tabular data, where each row corresponds to a record and each column is separated by a comma, making them easily readable and writable by both humans and machines.

See also:
https://r4ds.had.co.nz/data-import.html



Categorical data

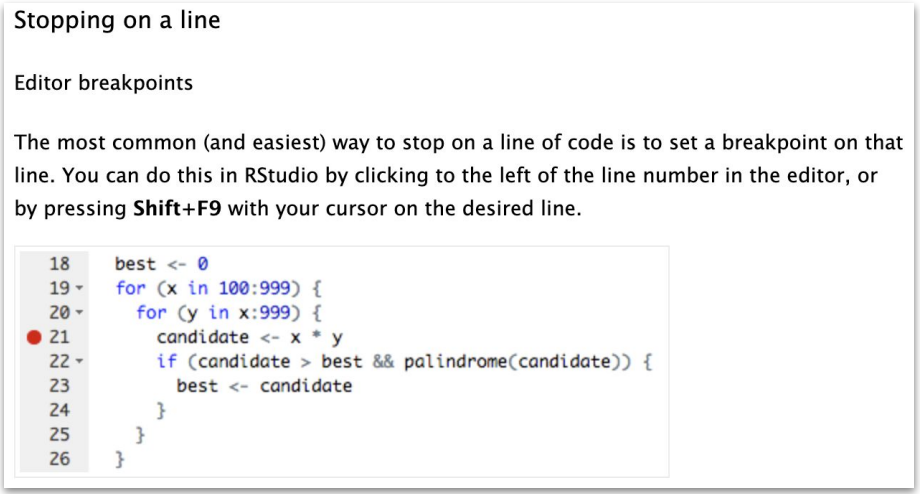
String / Character Data and categories/levels/factors in R

In R, "**strings as factors**" refers to converting character strings into factors when creating data frames, which is beneficial for categorical data analysis because factors store categorical variables with a fixed set of possible values (levels).

- Data that includes responses such as carnivore, omnivore, herbivore or things like survey data (yes, no, don't know) using factors is advantageous for statistical modeling and visualizations, as it allows R to treat these responses as categories rather than continuous values.
- Free-text responses or unique identifiers (like names or addresses) that don't have a predefined set of categories, it’s better not to convert them to factors, as this could lead to confusion and difficulties in data manipulation or text processing.

R Studio Debugging tools

- Debugging in RStudio
- Break points
- Running line by line
- print() statements
- Environment window / variables

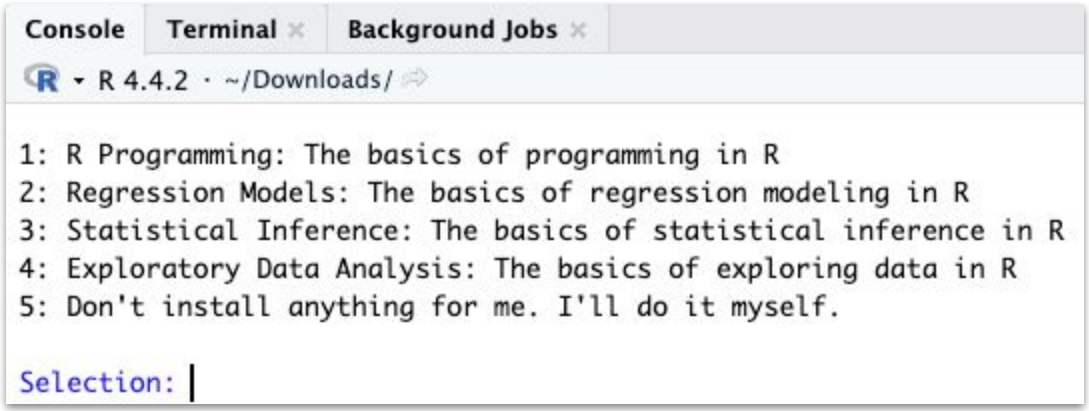


See also : https://support.posit.co/hc/en-us/articles/205612627-Debugging-with-the-RStudio-IDE

R swirl

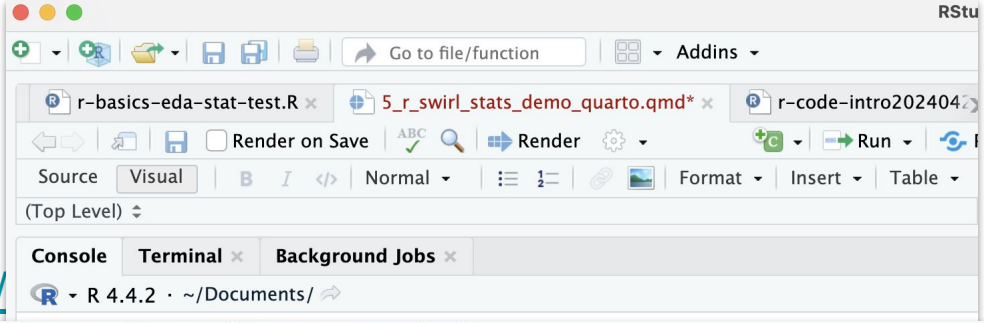
- What is R swirl?
- https://swirlstats.com/

```
install.packages("swirl")
library("swirl")
swirl()
```



Swirlstats

Swirlstats - https://




```
R 4.4.2 · ~/Documents/

> install.packages("swirl")
trying URL 'https://cran.rstudio.com/bin/macosx/big-sur-arm64/contr
5.tgz'
Content type 'application/x-gzip' length 352577 bytes (344 KB)
==================================================
downloaded 344 KB


The downloaded binary packages are in
        /var/folders/sg/cy13dc_j0vzg_r3xrvskyw500000gq/T//RtmpaTqWk
ges
> library(swirl)
```

Swirlstats

Swirlstats - https://swirlstats.com

```
| Let's get started!

...

| To begin, you must install a course. I can install a course for y
| internet, or I can send you to a web page
| (https://github.com/swirldev/swirl_courses) which will provide co
| and directions for installing courses yourself. (If you are not c
| the internet, type 0 to exit.)

1: R Programming: The basics of programming in R
2: Regression Models: The basics of regression modeling in R
3: Statistical Inference: The basics of statistical inference in R
4: Exploratory Data Analysis: The basics of exploring data in R
5: Don't install anything for me. I'll do it myself.
```

R swirl

- Swirl options - install course
- Swirl courses: https://github.com/swirldev/swirl_courses#swirl-courses

```
# make sure you have a recent version of swirl
install.packages("swirl")
library("swirl")
# install the course you want:
#install_course("course name here")
install_course()
swirl()
```

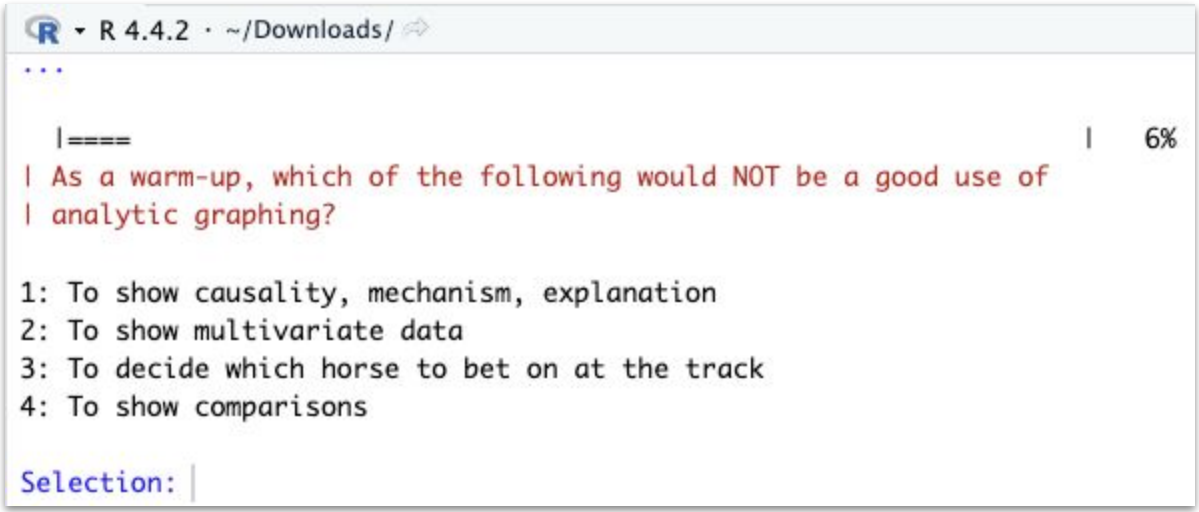
```
> install_course("R_Programming")
  |======================================================================| 100%

| Course installed successfully!

>
```

R swirl

- Swirl interactive questions
- Interactive coding
- Interactive readings

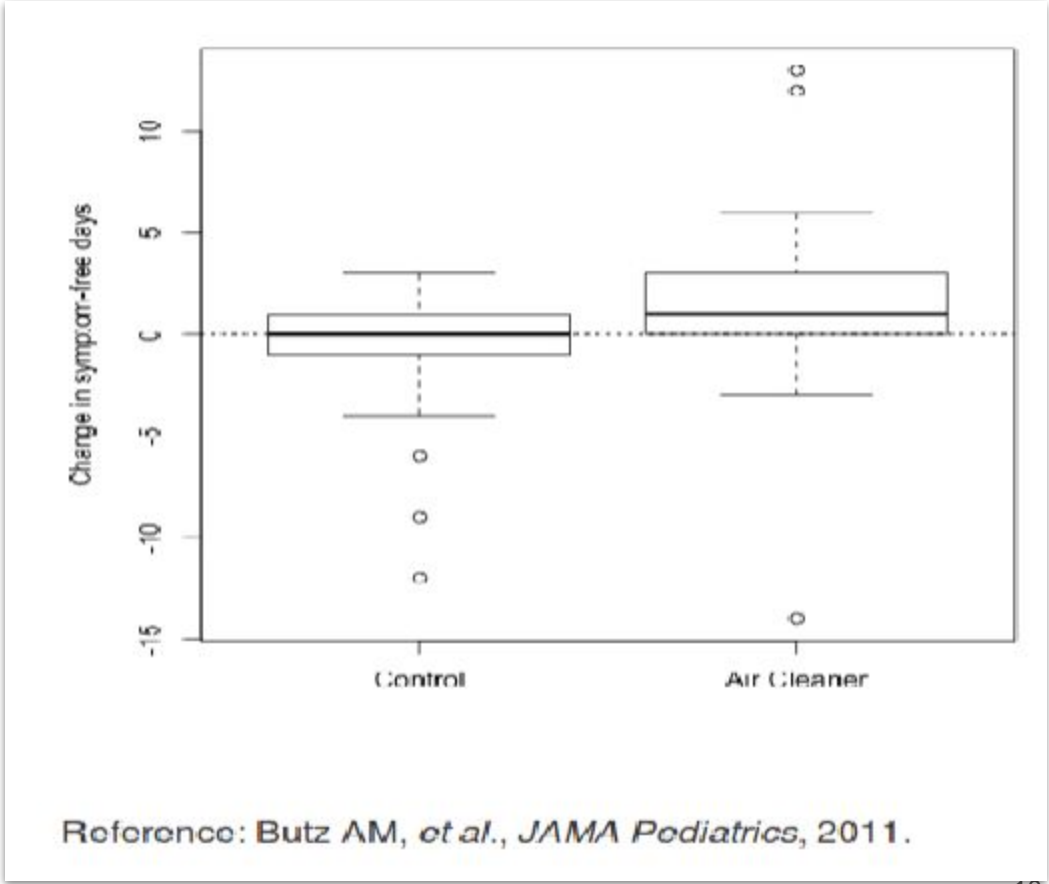


|================== | 28%
| What does this graph NOT show you?
1: Using the air cleaner makes asthmatic children sicker
2: 75% of the children using the air cleaner had at most 3 symptom-free days
3: Children in the control group had at most 3 symptom-free days
4: Half the children in the control group had no improvement

Selection: 1

| You're the best!

|==================== | 31%

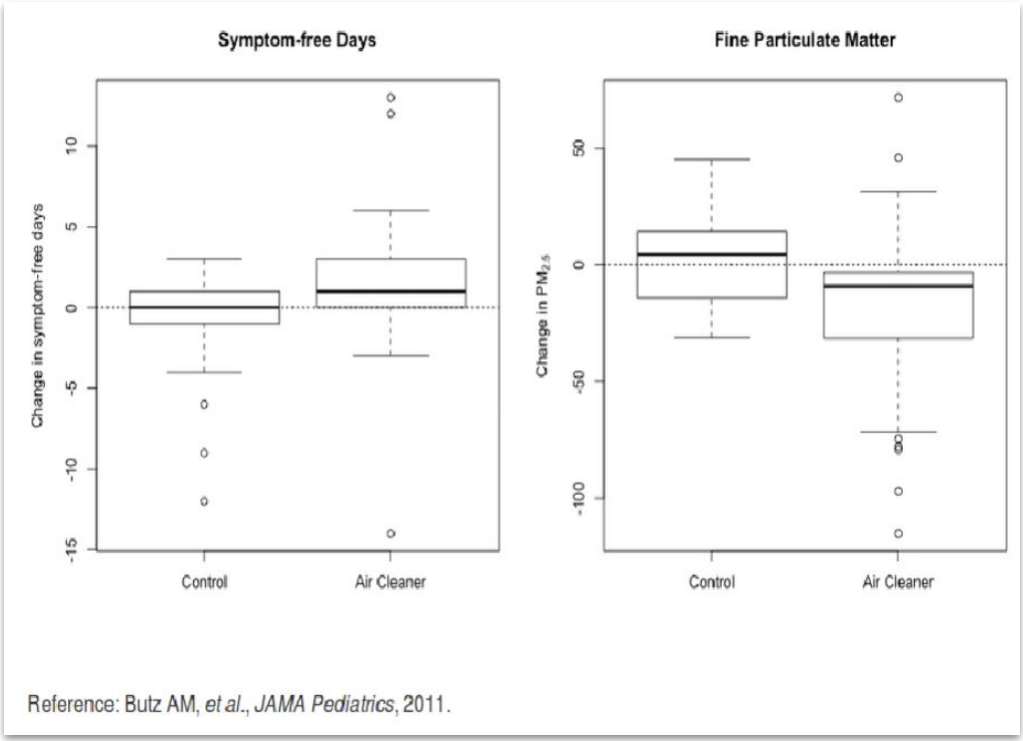


Swirlstats interactive: data, stats, visualizations and discussion points:

1: That the air in the control group is cleaner than the air in the other group
2: That the children in the control group are healthier
3: That the air cleaner increases pollution
4: That the air cleaner reduces pollution

Selection: 4

| Your dedication is inspiring!



|========================== | 39%

| By showing the two sets of boxplots side by side
| you're explaining your theory of why the air cleaner
| increases the number of symptom-free days. Onward!