Statistics for Linguists
08 July 2022

10:00	Workshop introduction
10:15	Loading and exploring datasets
10:45	Data transformation and coding
11:15	Practical exercise
12:15	Review of practical
12:30 - 13:30	LUNCH BREAK
13:30	lmer and glmer
14:30	Post-hoc analysis and model visualization
15:00	Practical exercise
16:00	Review of practical
16:15	Model building
17:00	End of workshop

Statistics for Linguists

Workshop introduction

Margreet Vogelzang – mv498@cam.ac.uk

About me

PhD in linguistics from the University of Groningen

PostDoc at University of Oldenburg, institute of Dutch studies

Current: PostDoc at University of Cambridge, Theoretical and Applied linguistics

Research interests

Language processing

Language acquisition

Cognitive science

Cognitive modeling

Neuroscience

Statistical modeling

About me

I learned statistics through courses (e.g. Summer School on Statistical Methods for Linguistics and Psychology https://vasishth.github.io/smlp2022/ and ZPID tidyverse workshop https://psycharchives.org/en/item/44bcacdf-f2bb-4891-b7e2-db33affb2dd8)

and by doing: the stats you use should fit your needs

Support team

- Alexander Cairncross
  PhD student in Theoretical and Applied Linguistics at the University of Cambridge
  Support during practical sessions, will answer questions

- Derya Nuhbalaoglu-Ayan
  GRADE Center Language
  Technical support

Statistics for Linguists
08 July 2022

*** There will be small breaks between the different topics!**

10:00	Workshop introduction
10:15	Loading and exploring datasets
10:45	Data transformation and coding
11:15	Practical exercise
12:15	Review of practical
12:30 - 13:30	LUNCH BREAK
13:30	lmer and glmer
14:30	Post-hoc analysis and model visualization
15:00	Practical exercise
16:00	Review of practical
16:15	Model building
17:00	End of workshop

Good morning!

What you can do now:

https://margreetvogelzang.github.io/

Contains the schedule, datasets and materials

Make sure you have R (R Studio) installed. It's free!

Overview

- This one-day course provides an introduction into statistics in R
- We will specifically discuss Mixed Models, which are a frequently used method in modern-day statistics

Learning objectives

- You will learn to load/import data
- Explore a dataset and create descriptive statistics
- Transform a dataset (if needed)
- Code your factors
- Build a mixed models
- Perform post-hoc statistics
- Visualize your data and your model

Notes

- I learned by doing, which means there may sometimes be more efficient ways to do something than what I show here

- There are multiple 'dialects' in R, such as tidyverse (https://www.tidyverse.org/). This requires slightly different syntax. You can use what you like, but I mostly use data tables.
  → There are too many ways to select variables:
  dfx, df"x", df[,"x"], df[[1]]

- Similarly, there are various approaches to building statistical models, and you may see different approaches used in articles in your field

Symbols and their names in R

Common operators

= - equal

. - dot

, - comma

> - greater than

< - less than

~ - twiddle

* - star

- - hyphen

_ - underscore

Quotation and comments

" - double quotation marks

' - single quotation marks

` - backticks

- hash

| - (vertical) bar

/ - (forward) slash

\ - backslash

Enclosures

( ) - parentheses

[ ]- (square) brackets

{ } - (curly) braces

<> - chevrons

R-specific operators

<- - assignment (left)

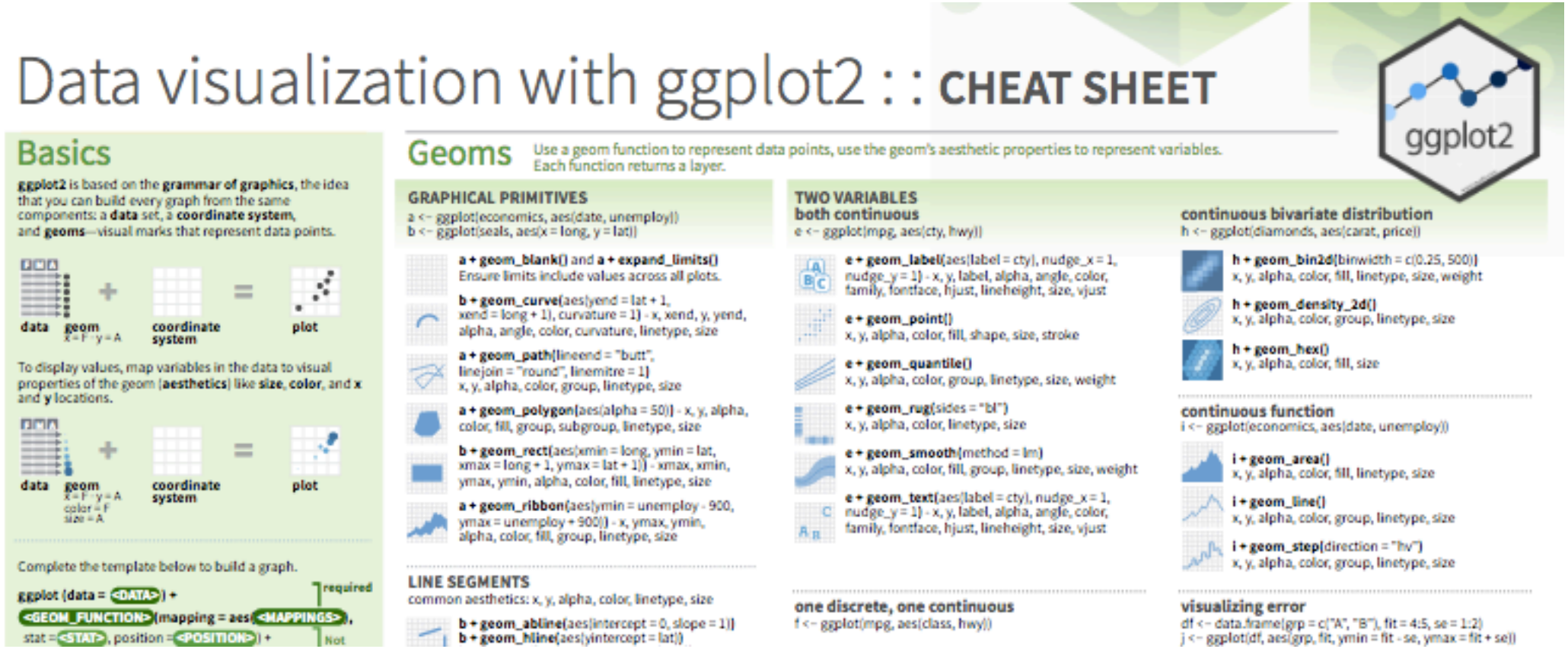
-> - right assignment

%>% - (magrittr) pipe

|> - (base) pipe

R cheat sheets: highly recommended!

https://www.rstudio.com/resources/cheatsheets/



Statistics for Linguists
08 July 2022

10:00	Workshop introduction
10:15	Loading and exploring datasets
10:45	Data transformation and coding
11:15	Practical exercise
12:15	Review of practical
12:30 - 13:30	LUNCH BREAK
13:30	lmer and glmer
14:30	Post-hoc analysis and model visualization
15:00	Practical exercise
16:00	Review of practical
16:15	Model building
17:00	End of workshop