.2 Plan design & data types

Instructions

Submit your answers to the seven first exercises through Campus Global. The remainder of your assignments concern self-studies.

Describing a study

Kanwal et al. (2017) taught subjects an artificial language with only three (made up) words: *zopudon*, *zopekil* and *zop*. The first two words refered to distinct objects (think: an apple for *zopudon* and a banana for *zopekil*). The short form *zop* could mean either. That is, *zop* is ambiguous, and thus can lead to misunderstandings. The core motivation for this study was to see if people would use the ambiguous word, *zop*, even though it is more risky than the unequivocal but longer alternatives *zopudon* and *zopekil* under certain experimental manipulations. Here's a glimpse of the data of the speakers:

```
library(dplyr)
library(readr)
read_csv('senderdata.csv') %>%
  glimpse()
```

```
## Rows: 1,280
## Columns: 5
## $ pairnum <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ IP      <chr> "67.85.42.18", "67.85.42.18", "67.85.42.18", "67.85.42.18", "6~
## $ trial   <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,~
## $ display <dbl> 0, 3, 0, 0, 2, 1, 1, 0, 0, 0, 2, 1, 3, 3, 1, 1, 1, 2, 3, 0, 1,~
## $ label   <chr> "zop", "zopudon", "zop", "zopekil", "zopudon", "zopekil", "zop~
```

The column `pairnum` identifies each pair of subjects (one is the speaker, the other the receiver); the `IP` column identifies each subject; `trials` keep track of the order in which the trials happened; `display` codes whether one object (0/1) or another (2/3) was displayed to the subject; and `label` shows what the speaker actually said to communicate this object.

1. What kind of study is this? Observational, experimental, or simulation?
2. What kind of variable is `pairnum`?
3. What kind of variable is `trial`?
4. What kind of variable is `label`?s

Inspect the sender data ('senderdata.csv') from Kanwal et al. yourself.

5. How many times did the sender with `IP` 67.85.42.18 say *zop*?
6. How many unique pairs participated in the experiment?
7. How many sender trials did the experiment have for each subject?s
8. Is this data *tidy*; *untidy* or *almost tidy*?

Self-study

1. Work through Chapter 3 and Chapter 4 of Introduction to Data Analysis (Franke 2021): *Data, variables & experimental designs*;
2. If you haven't already, make sure you have a working R environment: either locally –on your computer– or by having made sure you can execute code in a colab document
3. Apply the terminology from this session to your analysis question of interest. Change your research question if you have come across another that you find more interesting # References

Franke, Michael. 2021. *An Introduction to Data Analysis*.

Kanwal, Jasmeen, Kenny Smith, Jennifer Culbertson, and Simon Kirby. 2017. "Zipf's Law of Abbreviation and the Principle of Least Effort: Language Users Optimise a Miniature Lexicon for Efficient Communication." *Cognition* 165. Elsevier BV: 45–52. https://doi.org/10.1016/j.cognition.2017.05.001.