Confidence intervals and data wrangling

21 February 2020

Modern Research Methods

Assignment 5

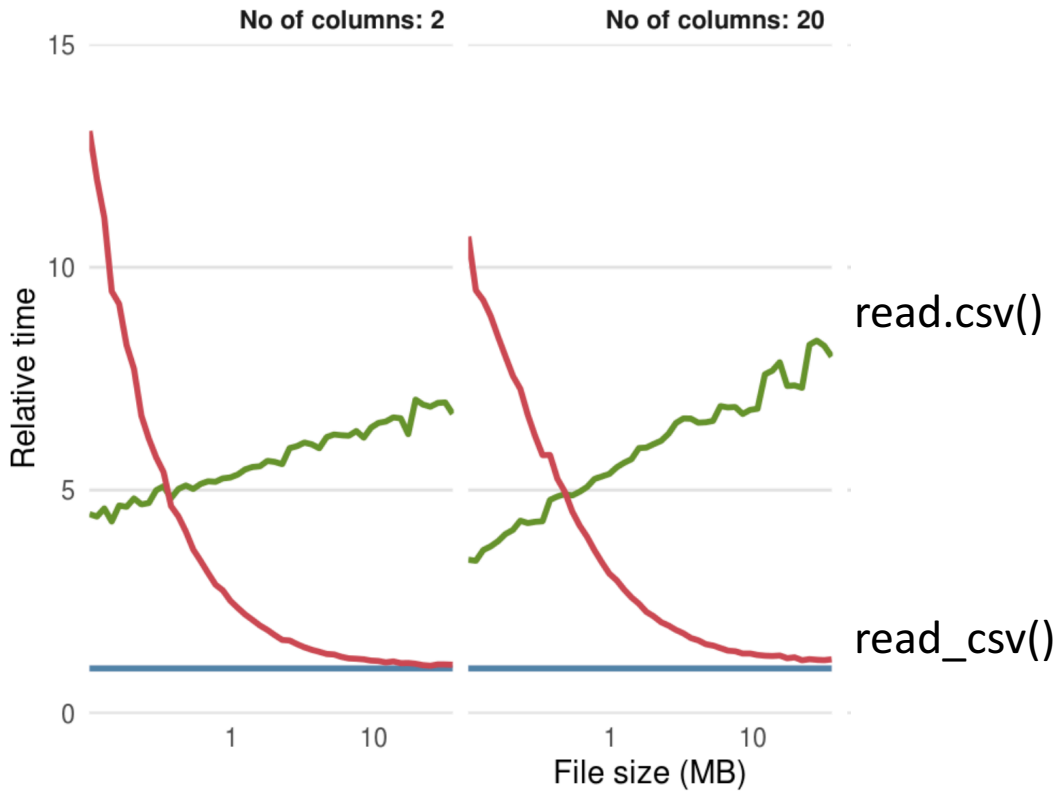
- Last assignment before midterm
- Confidence intervals and data wrangling
- Data concerns project replicating an experiment.

- In lab today:
  - A few points of confusion
  - Introduce Assignment 5 dataset.
  - Data wrangling primer.

Some points of confusion

1. Use read_csv() rather than read.csv()
2. geom_line() vs. geom_smooth()
3. Indices start at 1 (not zero)
4. kable() – prints a data frame nicely
5. Space after # to get it to format correctly.

Good 10 minute tutorial for markdown: https://www.markdowntutorial.com/



source: https://csgillespie.github.io/

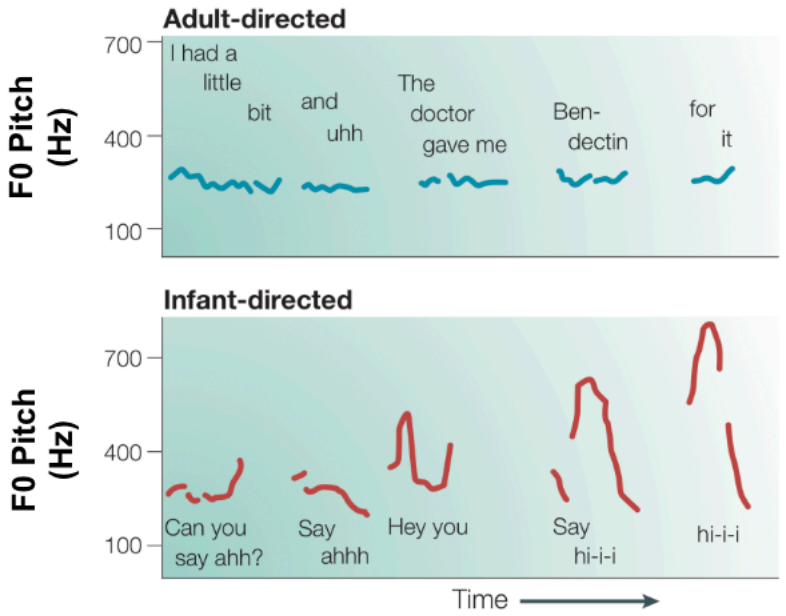
	Original	Reproduction	Replication
Population			
Question			
Hypothesis			
Exp. Design			
Experimenter			
Data	01100 10110 11110	01100 10110 11110	01100 10110 11110
Analyst			
Code			
Estimate			
Claim			

Original:

Preference for Infant-directed Speech in the First Month after Birth

Robin Panneton Cooper
Virginia Polytechnic Institute and State University

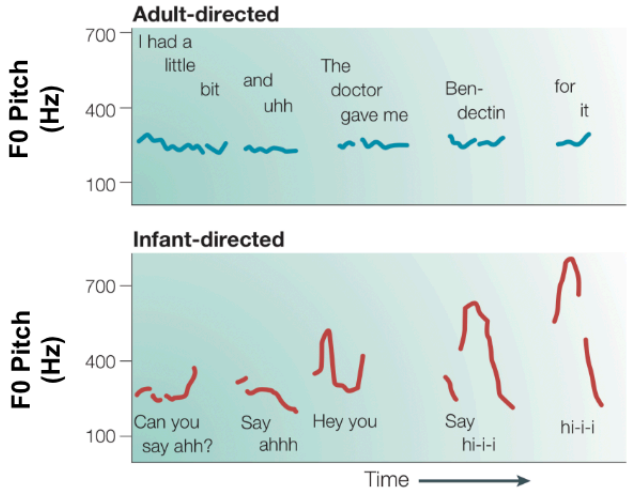
Richard N. Aslin
University of Rochester

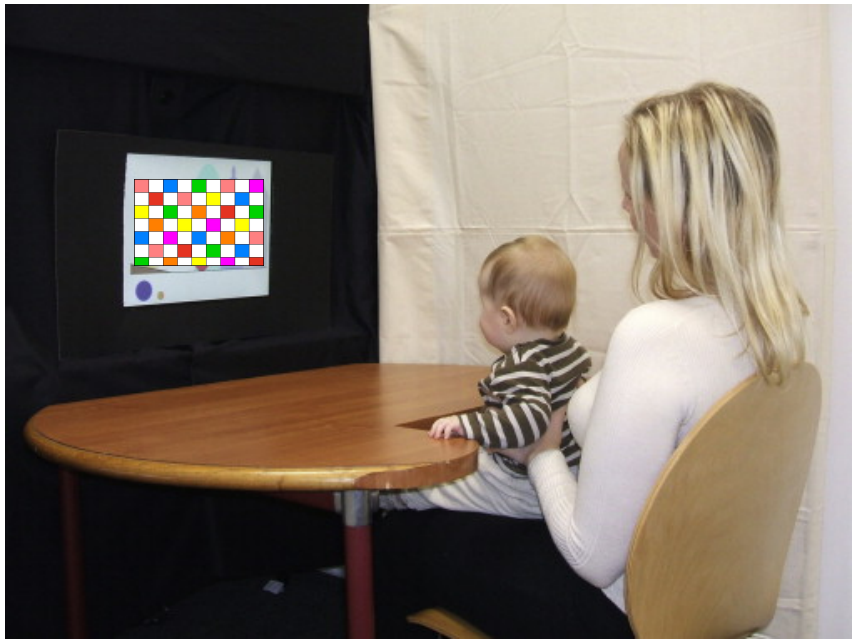


Shorter utterances,
Higher, varied pitch,
Longer pauses

Cooper & Aslin (1990)

Do infants prefer IDS to ADS?





Source: Moll & Tomasello, 2010

Dependent measure: Looking time to checkerboard

Independent variable: ADS vs. IDS played in pairs of trials within subjects

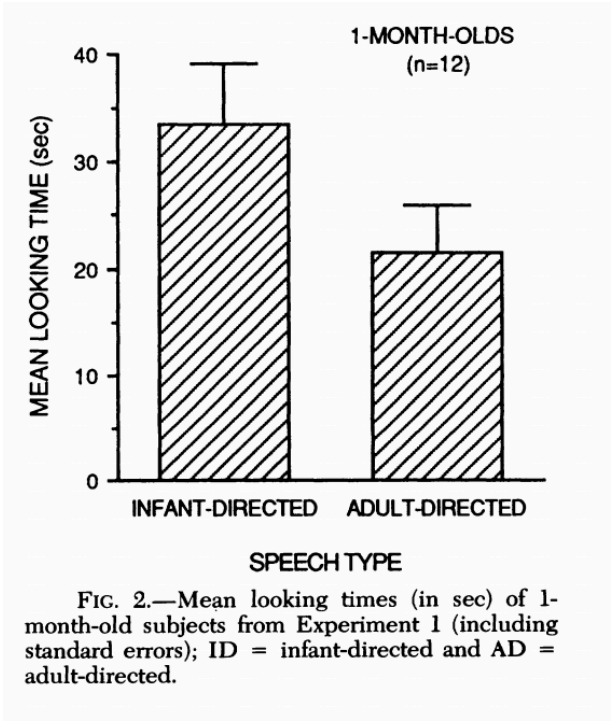
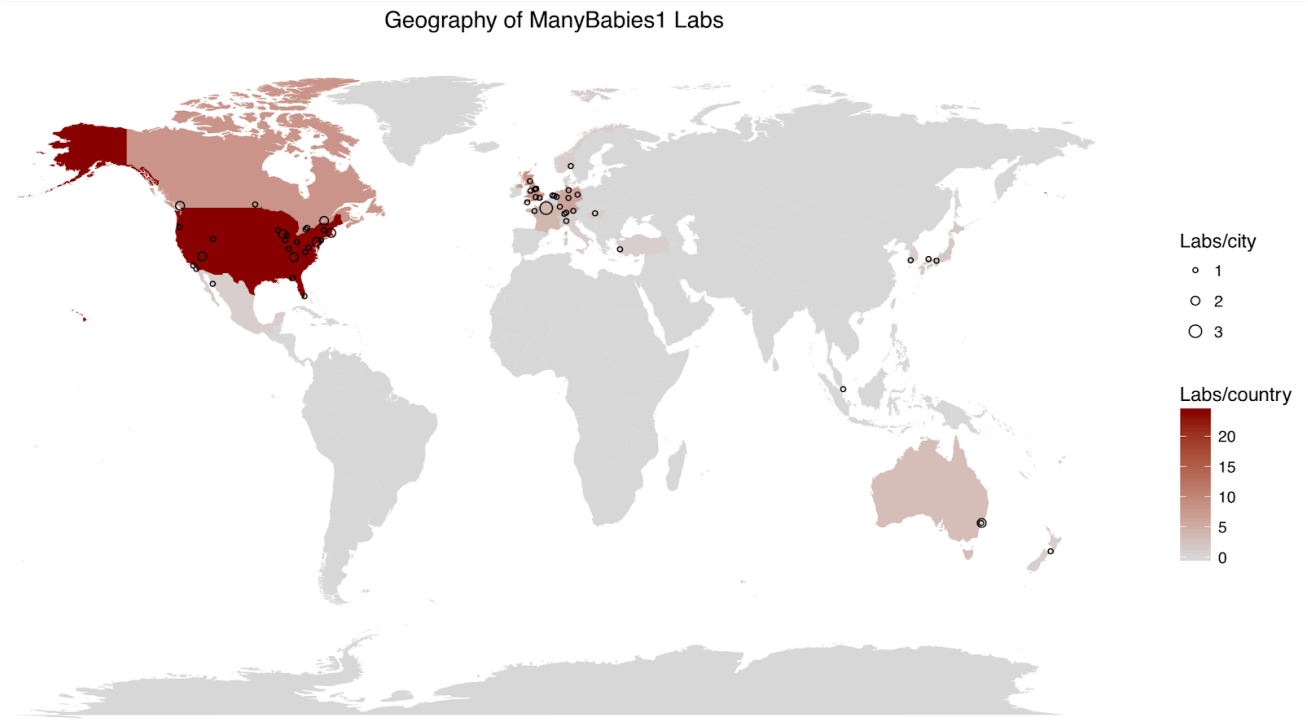


FIG. 2.—Mean looking times (in sec) of 1-month-old subjects from Experiment 1 (including standard errors); ID = infant-directed and AD = adult-directed.

ManyBabies, 2017

- Multi-lab effort to replicate this effect (paper)
- Each lab conducted their own replication of Cooper & Aslin (1990), with standardization of the paradigm across labs



68 labs, 2773 babies!

Assignment 5 includes data from 6 labs.

Data wrangling

- Tidy data ("long format")
- What if your data are not tidy?
- Two functions in tidyverse (tidyr package) for reshaping data
  - spread() -> long to wide
  - gather() -> wide to long
- In the process of renaming these functions:
  - spread -> pivot_wider()
  - gather -> pivot_longer()

https://rstudio.cloud/learn/primers/4.1

Artwork by @allison_horst

Tidy Your Data

Reshape Data

- Welcome
- Tidy Data
- Gathering columns
- Spreading columns

Start Over

Welcome

The tools that you learned in the previous Primers work best when your data is a specific way. This format is known as **tidy data** and it appears throughout the You will spend a lot of time as a data scientist wrangling your data into a useab is important to learn how to do this fast.

This tutorial will teach you how to recognize tidy data, as well as how to reshap into a tidy format. In it, you will learn the core data wrangling functions for the ti

- `gather()` - which reshapes wide data into long data, and
- `select()` - which reshapes long data into wide data

Typo!! ------->

This tutorial uses the core tidyverse packages, including ggplot2, dplyr, and tid the `babynames` package. All of these packages have been pre-installed and p your convenience.

Should say spread()

Click the Next Topic button to begin.

Next Topic

https://rstudio.cloud/learn/primers/4.1