Confidence intervals and replicability

8 October 2021

Modern Research Methods

Feedback on Roderick's guest lecture

- Survey anonymous, voluntary, and confidential
- Administered by the Eberly Center
- Please complete by October 10

Lab today

- Common errors from previous assignment
- Introduce Assignment 5 dataset
- Make class study guide for confidence intervals

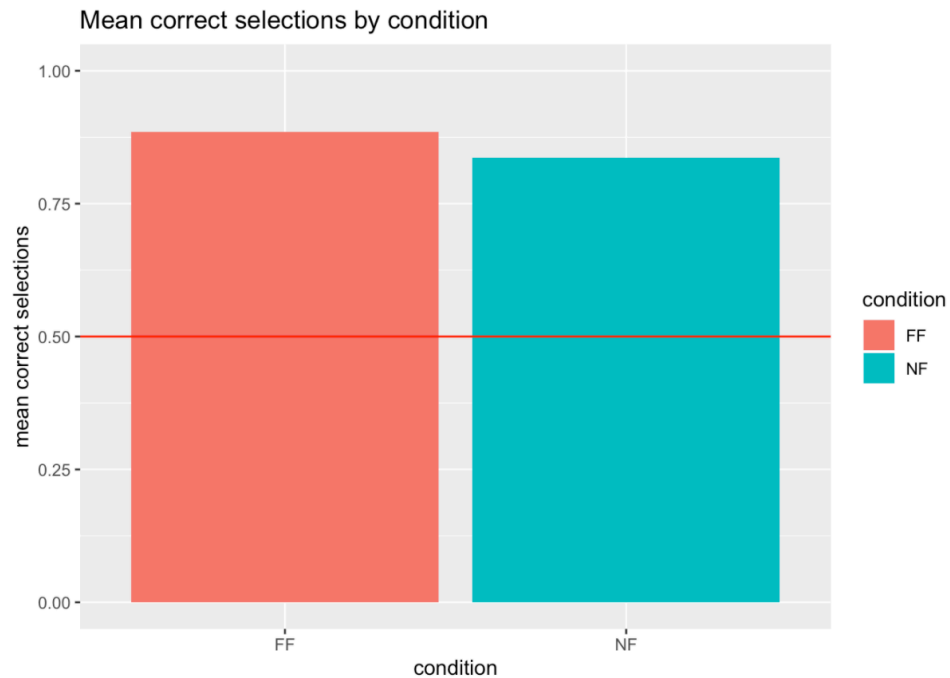
Exercise 5 Use the `subject_means` data frame to calculate the mean correct by condition. Plot the result as a bar plot. Include the following things:

- a y-axis that scales from 0 to 1 (use `ylim`).
- each condition as a different fill
- an appropriate title
- appropriate x- and y-axis labels
- a red horizontal line indicating chance performance (use `geom_hline()`; geom_hline takes one parameter, yintercept).

Which condition are children better at?

```
overall_means <- subject_means %>%
  group_by(condition) %>%
  summarize(mean_correct = mean(prop_correct))
```

```
ggplot(overall_means, aes(x = condition, y = mean_correct, fill = condition)) +
  ggtitle("Mean correct selections by condition") +
   geom_bar(stat = "identity") +
  ylab("mean correct selections") +
  ylim(0,1) +
  geom_hline(yintercept = .5, color = "red")
```



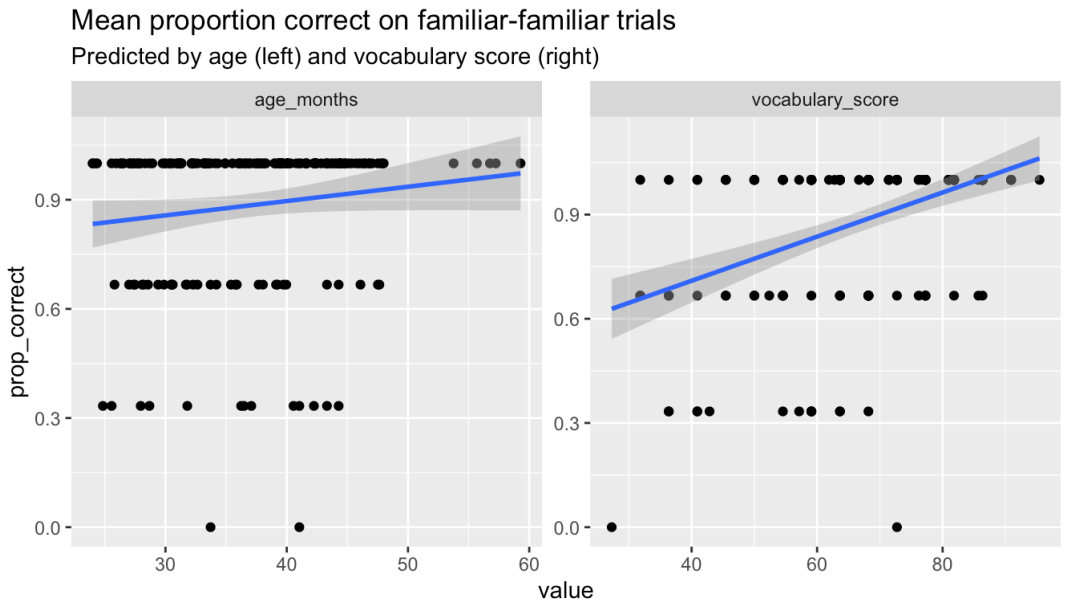
Exercise 12 Recreate the plot below, where each point corresponds to an individual child.

```
subject_means_with_years_months <- me_data %>%
  filter(condition == "FF")  %>%
  group_by(sub_id, age_months, vocabulary_score) %>%
  summarize(prop_correct = sum(correct)/n())
```

```
## `summarise()` has grouped output by 'sub_id', 'age_months'. You can over
ent.
```

```
data_for_faceting <- pivot_longer(subject_means_with_years_months,cols = 2:

ggplot(data_for_faceting, aes(x=value, y = prop_correct)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~name, scales = "free") +
  labs(title = "Mean proportion correct on familiar-familiar trials",
       subtitle =  "Predicted by age (left) and vocabulary score (right)")
```




```
head(subject_means_with_years_months)
```

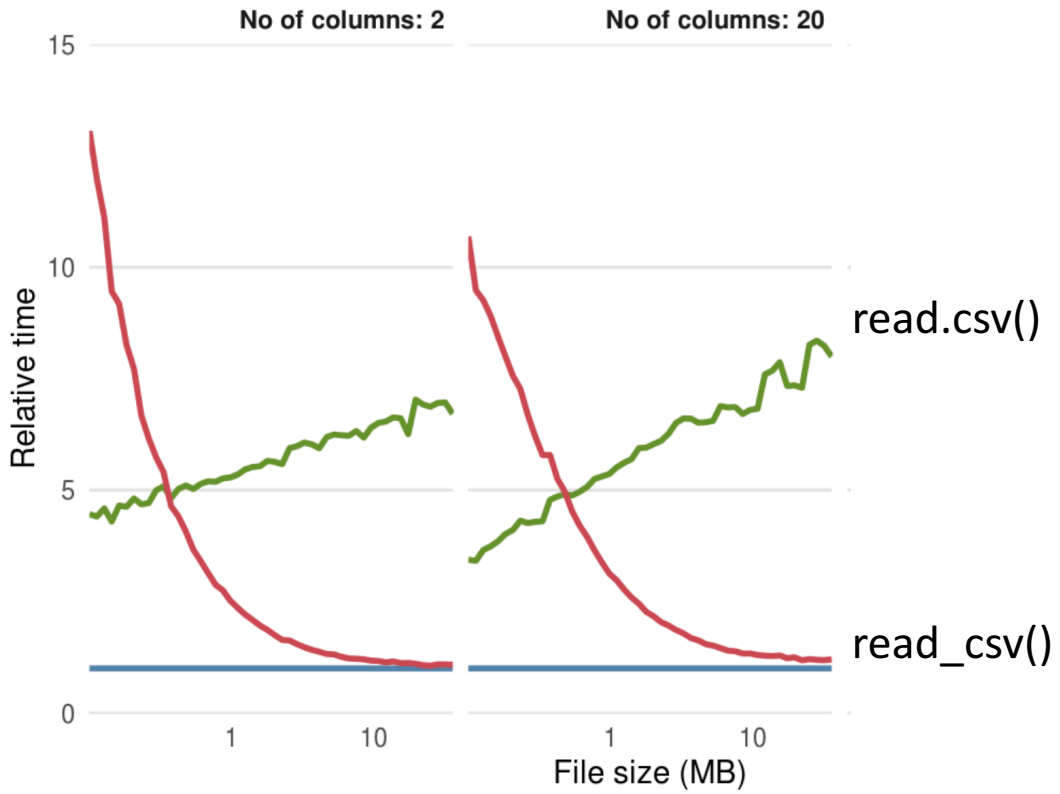
```
## # A tibble: 6 × 4
## # Groups:   sub_id, age_months [6]
##   sub_id age_months vocabulary_score prop_correct
##    <dbl>      <dbl>            <dbl>        <dbl>
## 1      1       47.0             72.7        1
## 2      2       30.4             59.1        1
## 3      3       41.6             77.3        1
## 4      4       41.5             72.7        1
## 5      5       42.4             68.2        1
## 6      6       32.3             63.6        0.667
```

```
head(data_for_faceting)
```

```
## # A tibble: 6 × 4
## # Groups:   sub_id [3]
##   sub_id prop_correct name             value
##    <dbl>        <dbl> <chr>            <dbl>
## 1      1            1 age_months        47.0
## 2      1            1 vocabulary_score  72.7
## 3      2            1 age_months        30.4
## 4      2            1 vocabulary_score  59.1
## 5      3            1 age_months        41.6
## 6      3            1 vocabulary_score  77.3
```

Some other points of confusion

1. Use read_csv() rather than read.csv()
2. Headers
   - Space after # to correct format
   - Need blank line before so doesn't show up next to plot
   - Good 10 minute tutorial for markdown: https://www.markdowntutorial.com/
3. Geom_histogram, binwidth, no stat= "identity"



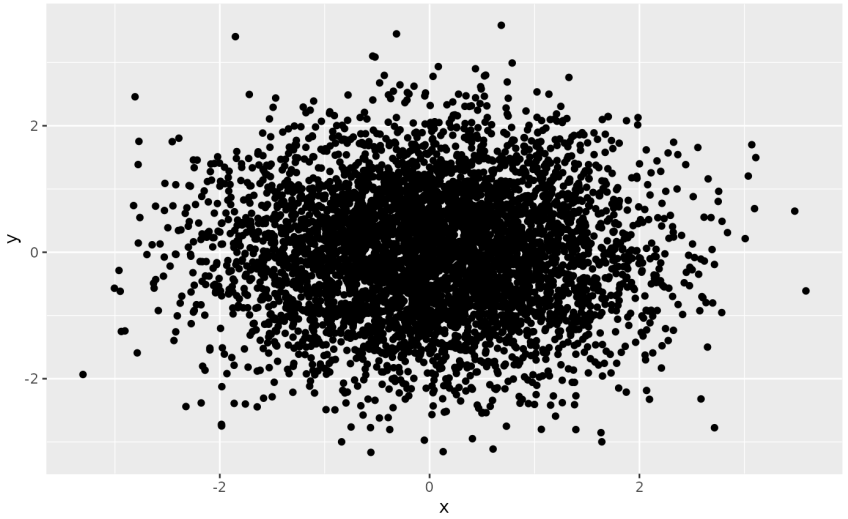
source: https://csgillespie.github.io/

Assignment 5

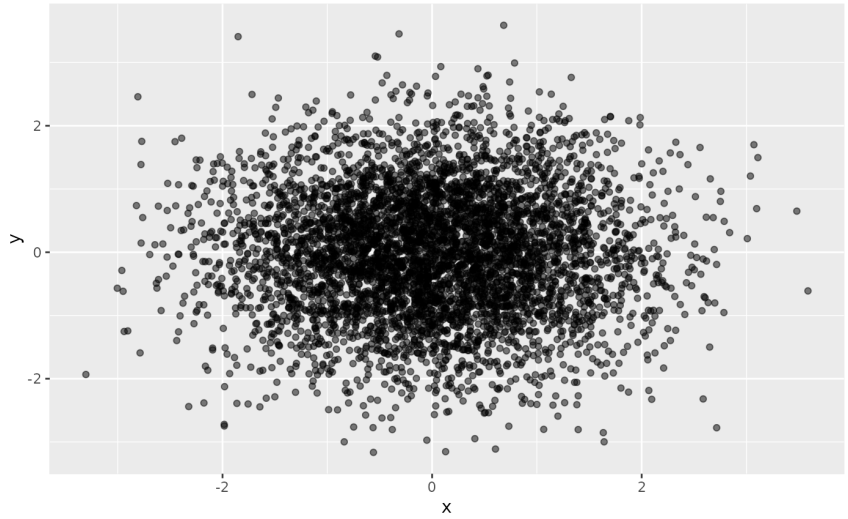
- Last assignment before midterm
- Focuses on estimating means from experimental data and quantifying uncertainty

New aesthetic: alpha

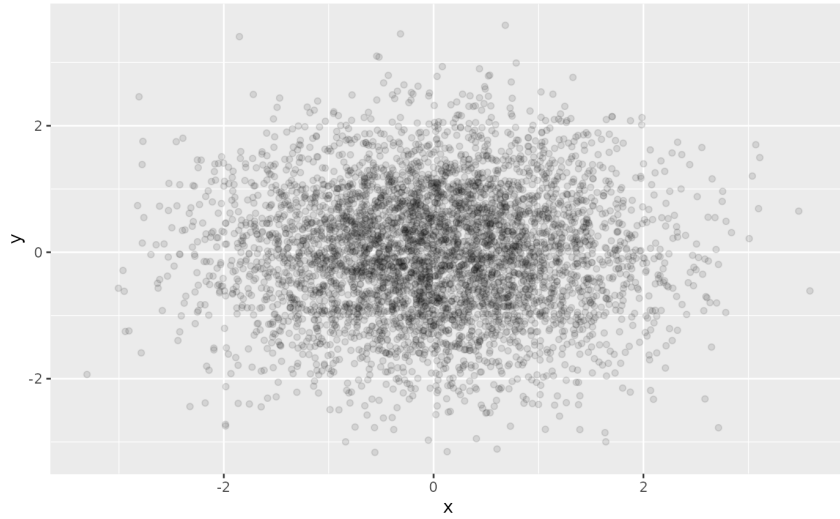
- Geom opacity
- Range from 0 to 1, with lower values corresponding to more transparent colors



geom_point()



geom_point(alpha = .5)



geom_point(alpha = .1)

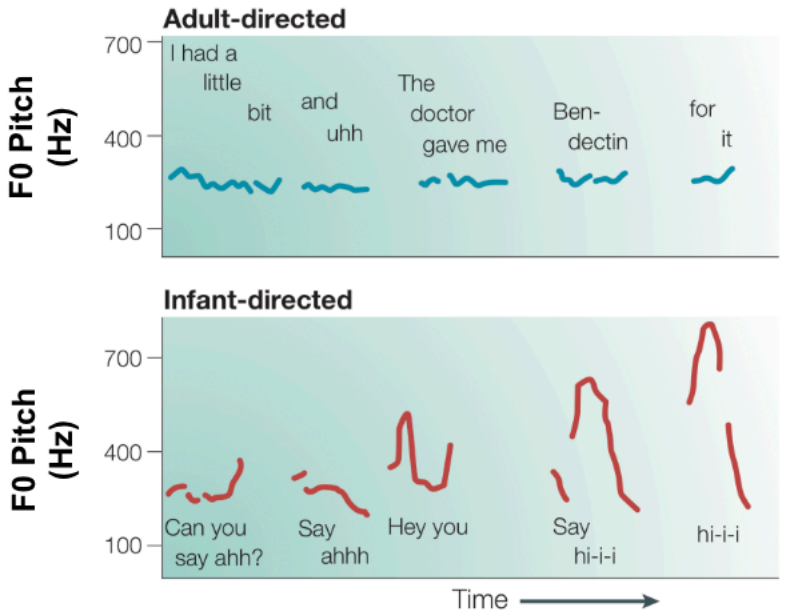
	Original	Reproduction	Replication
Population			
Question			
Hypothesis			
Exp. Design			
Experimenter			
Data	01100 10110 11110	01100 10110 11110	01100 10110 11110
Analyst			
Code			
Estimate			
Claim			

Original:

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

Robin Panneton Cooper
Virginia Polytechnic Institute and State University

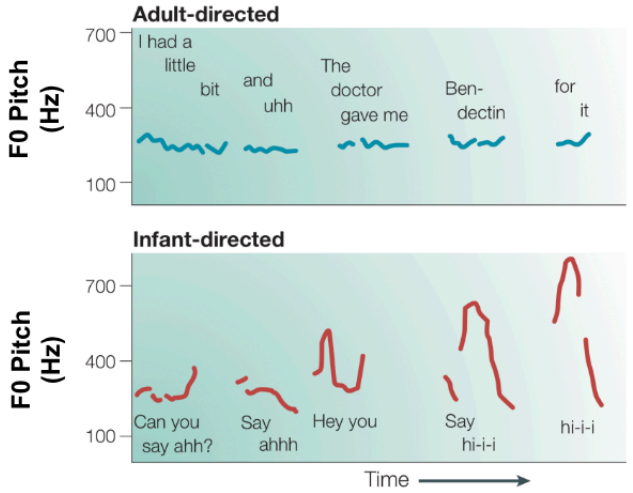
Richard N. Aslin
University of Rochester

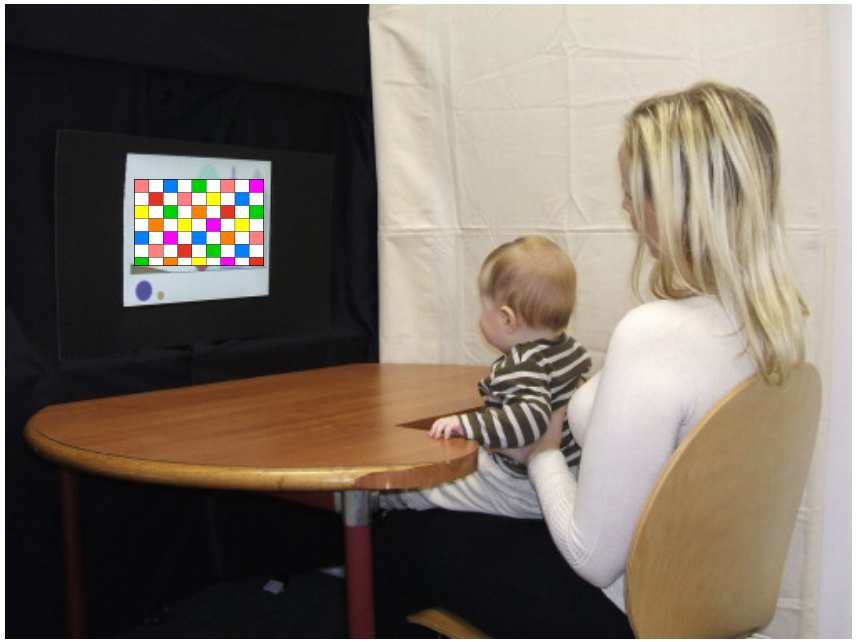


Shorter utterances,
Higher, varied pitch,
Longer pauses

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

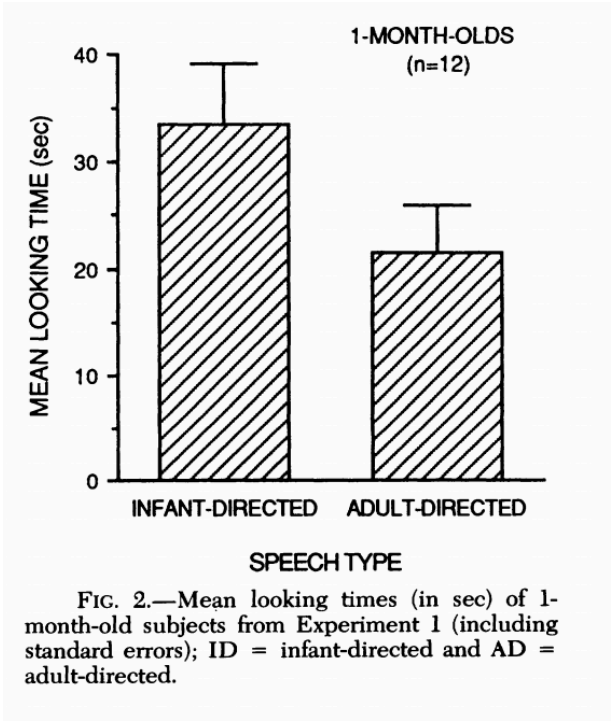
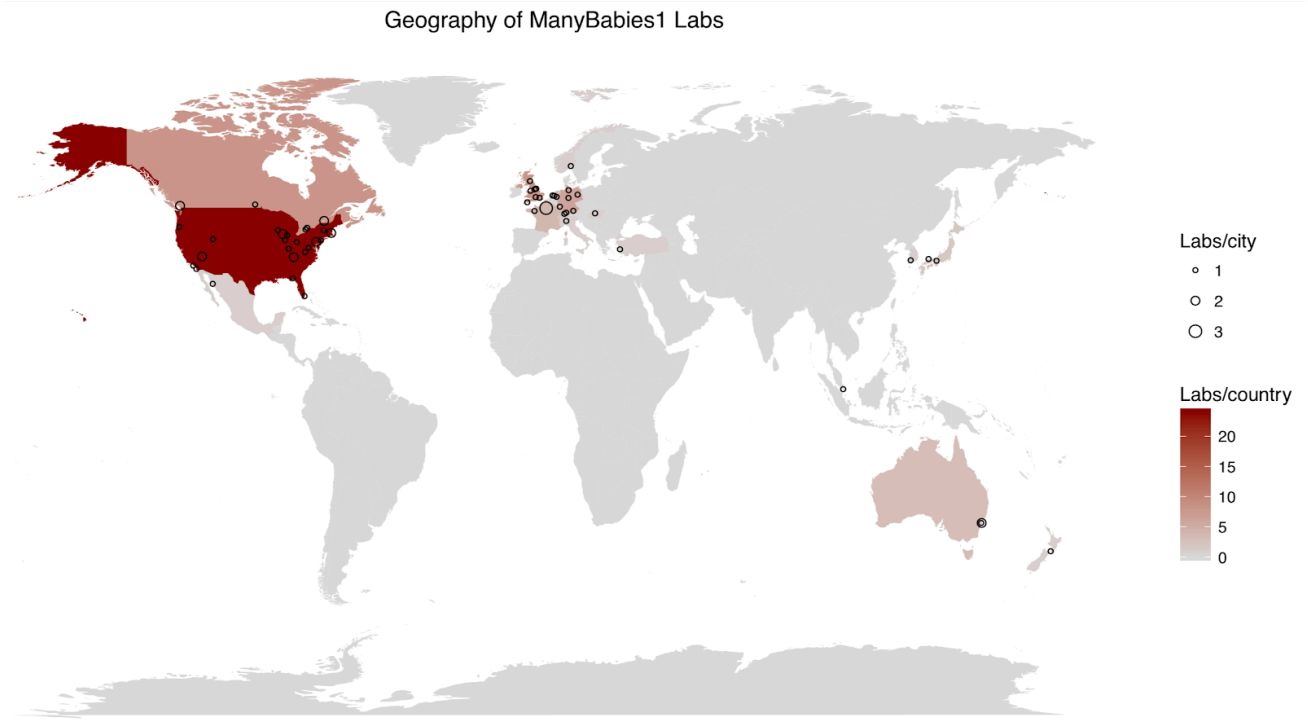


FIG. 2.—Mean looking times (in sec) of 1-month-old subjects from Experiment 1 (including standard errors); ID = infant-directed and AD = adult-directed.

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.

Class CI study guide