

# Intro to Data Wrangling

Joscelin Rocha-Hidalgo

she/her/hers

@JoscelinRocha

**Slides adapted from David Keyes (@dgkeyes), inspired by Danielle Navarro  
(@djnavarro) and Paul Campbell (@paulcampbell91)**

# Child Health and Development Studies (CHDS)

"Birth weight, date, and gestational period collected as part of the Child Health and Development Studies in 1961 and 1962. Information about the baby's parents – age, education, height, weight, and whether the mother smoked is also recorded."



- Website: <https://www.stat.berkeley.edu/users/statlabs/papers/sample.pdf>
- R package:  
<https://vincentarelbundock.github.io/Rdatasets/doc/mosaicData/Gestation.html>



# Tidyverse



# The Pipe

I would read each pipe as "then." For example:

```
data %>%
  filter(age < 25) %>%
  group_by(ed) %>%
  summarize(mean_gestation = mean(gestation, na.rm = TRUE))
```

Art by @allison\_horst



# Shortcuts



Windows

control-shift-M

# Shortcuts



Windows

control-shift-M



Mac

command-shift-M

- These are the functions we will go over:

1. rename
2. clean\_names
3. toupper/tolower
4. separate/unite
5. select
6. filter
7. mutate
8. case\_when
9. summarize
10. group\_by
11. relocate
12. pivot\_longer/pivot\_wider