

# Analysis of CSV File Reading in R

Quang Nguyen\* and Robert Tedesco\*

\* Loyola University Chicago

STAT 407 Fall 2021

# Outline

- ▶ Background
- ▶ Design and Data
- ▶ ANOVA
- ▶ Additional Analysis
- ▶ Discussion

# Background

- ▶ Statisticians struggle with reading in large dataset (Matthews, 2021)
- ▶ `read.csv`, part of base R (R Core Team, 2021), is old and slow.
- ▶ Old dogs need to learn new tricks
- ▶ There are better and more efficient data I/O algorithms in R (Gillespie & Lovelace, 2021)
  - ▶ `rio::import` (Chan et al., 2021)
  - ▶ `readr::read_csv` (Wickham & Hester, 2021) - part of `tidyverse` (Wickham et al., 2019)
  - ▶ `data.table::fread` (Dowle & Srinivasan, 2021)

# Objective

Our goal is to design an experiment. . .

- ▶  $3^k$  factorial design (Montgomery, 2012)
- ▶  $n = 20$  replicates
- ▶ How were the data generated?

# ANOVA

- ▶ Insert table
- ▶ Post-hoc

# Additional Analysis

► don't know yet

- Take advantage of more efficient algorithms.



# Cheers.

- ▶ S/o out Drs. Perry and Matthews
- ▶ Greg's YouTube video: [youtu.be/E5KJkooW4RY](https://youtu.be/E5KJkooW4RY)
- ▶ GitHub: [github.com/qntkhvn/read\\_speed](https://github.com/qntkhvn/read_speed)
- ▶ Question?

# References I

- Chan, C., Chan, G. C., Leeper, T. J., & Becker, J. (2021). *Rio: A swiss-army knife for data file i/o*.  
<https://CRAN.R-project.org/package=rio>
- Dowle, M., & Srinivasan, A. (2021). *Data.table: Extension of 'data.frame'*.  
<https://CRAN.R-project.org/package=data.table>
- Gillespie, C., & Lovelace, R. (2021). *Efficient r programming*.  
<https://csgillespie.github.io/efficientR>
- Matthews, G. (2021). *Old dog (statistician) learns new trick (read\_csv)*.  
<https://youtu.be/E5KJkooW4RY>.
- Montgomery, D. C. (2012). *Design and analysis of experiments, 8th edition*. John Wiley & Sons.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.  
<https://www.R-project.org/>

# References II

- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.  
<https://doi.org/10.21105/joss.01686>
- Wickham, H., & Hester, J. (2021). *Readr: Read rectangular text data*.  
<https://CRAN.R-project.org/package=readr>