# Statistical inference via data science: A "tidy" approach

**Albert Y. Kim**
**@rudeboybert**

**Joint Math Meetings**
**Denver CO USA**
**January 18, 2020**


Statistical inference **via** data science…

# Context: Definitions

Intro stats

- Course where students either will or have the possibility of being producers of statistics
- Not Gen Ed & consumers of statistics

Data science

- Set of tools

# What is the tidyverse?


## R packages for data science

The tidyverse is an opinionated **collection of R packages** designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

From: [tidyverse.org](tidyverse.org)

# Why use the tidyverse?

1. It bridges the gap between tools for *learning* statistics & tools for *doing* statistics

tidy tools manifesto principle #4: Design for humans

**David Robinson**

*Data Scientist at Stack Overflow, works in R and Python.*

## Teach the tidyverse to beginners

A few years ago, I wrote a post Don't teach built-in plotting to beginners (teach ggplot2). I argued that ggplot2 was not an advanced approach meant for experts, but rather a suitable introduction to data visualization.

> *Many teachers suggest I'm overestimating their students: "No, see, my students are beginners...". If I push the point, they might insist I'm not understanding just how much of a beginner these students are, and emphasize they're looking to keep it simple and teach the basics, and that that students can get to the advanced methods later....*

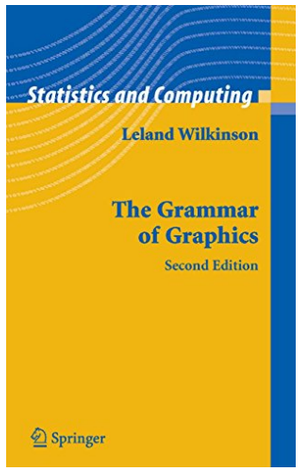
GitHub

stackoverflow

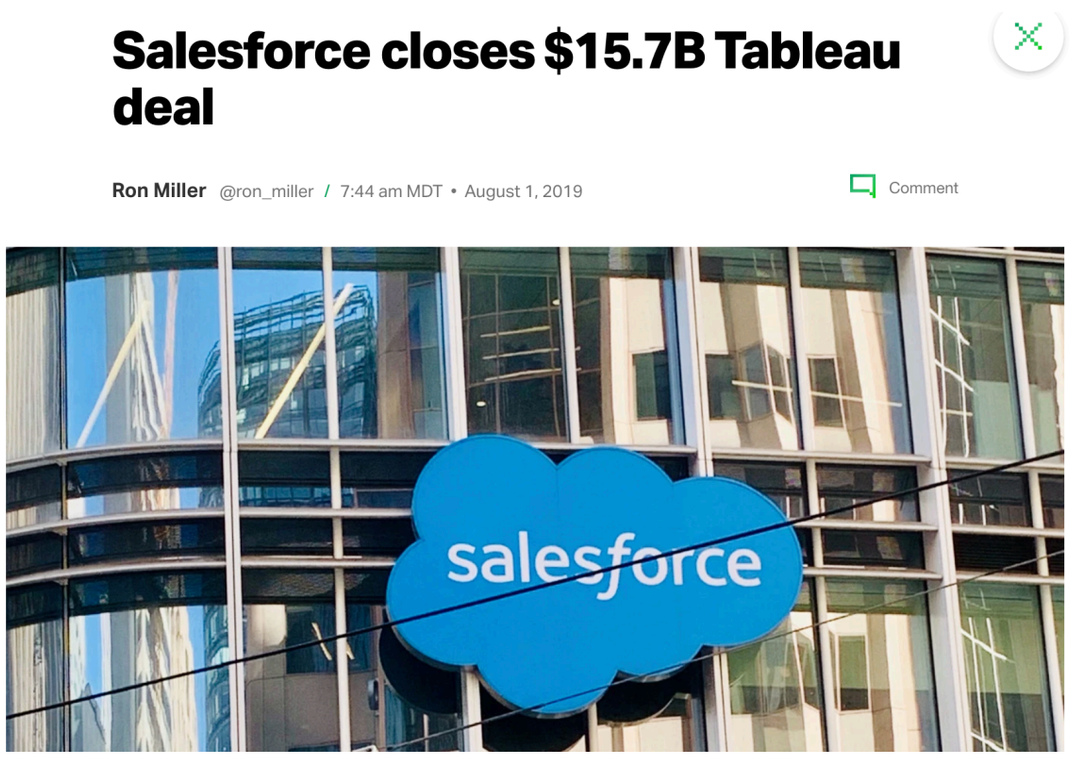
From: A.McNamara (2015), D.Robinson blogpost, tidy tools manifesto

# Why use the tidyverse?

## 2.a) It's transferable: Data visualization






From: Wilkinson (2005), ggplot2 package, TechCrunch

# Why use the tidyverse?

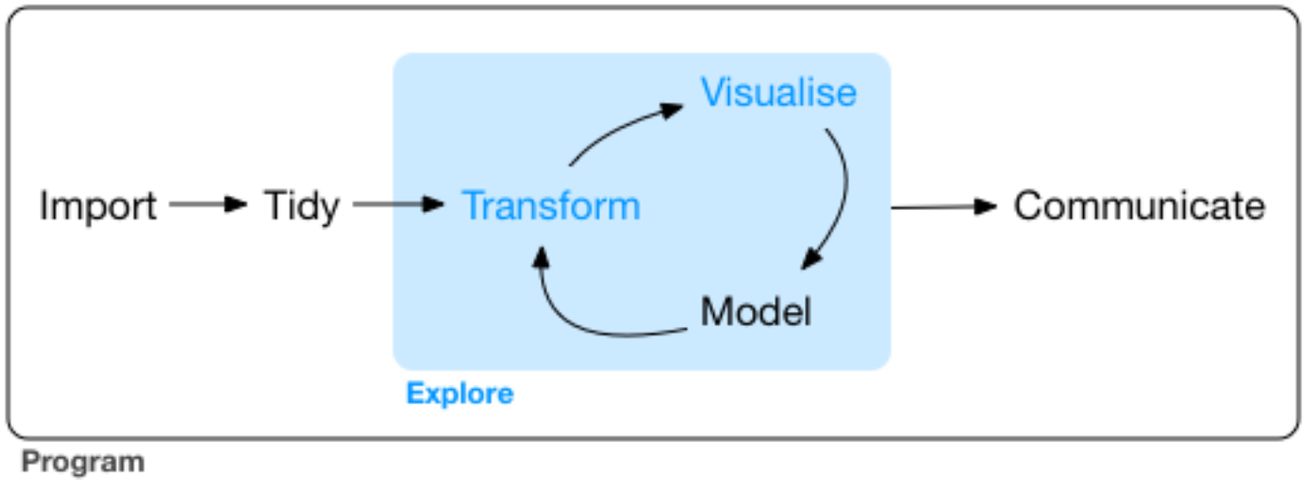
## 2.b) It's transferable: Data wrangling



From: Wilkinson (2005), Database Normalization

# Why use the tidyverse?

## 3. Encourages students to “play the whole game”


- Exploratory data analysis
- To do NO data wrangling betrays true nature of work

From: YouTube, r4ds (2017),

# In Practice: Intro Stats

Statistical Modeling

Statistical inference

# Statistical Modeling

Starbucks

# Statistical Modeling

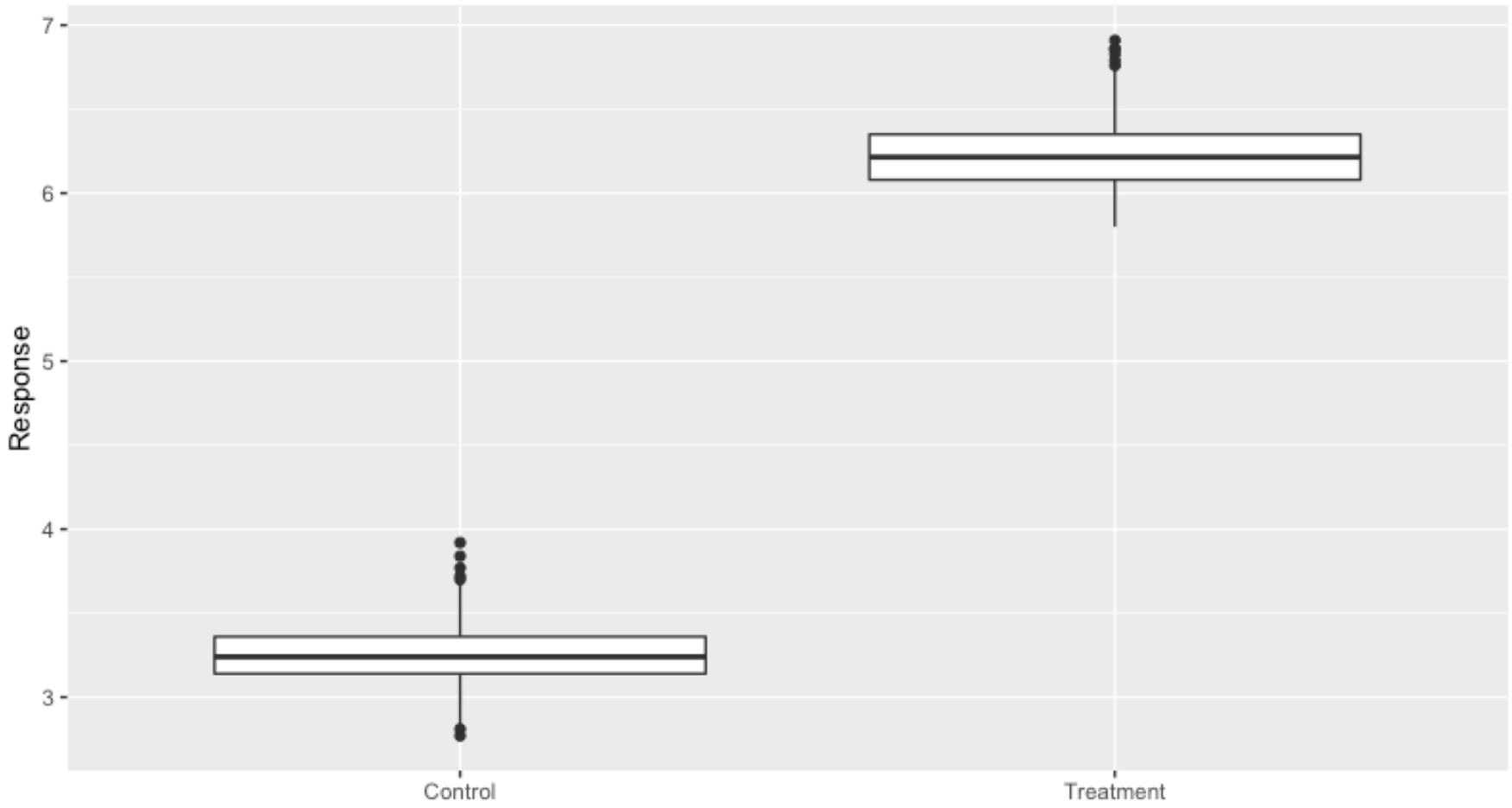
Model selection: interaction model

# Statistical Modeling

Model selection: parallel slopes
Occam's razor

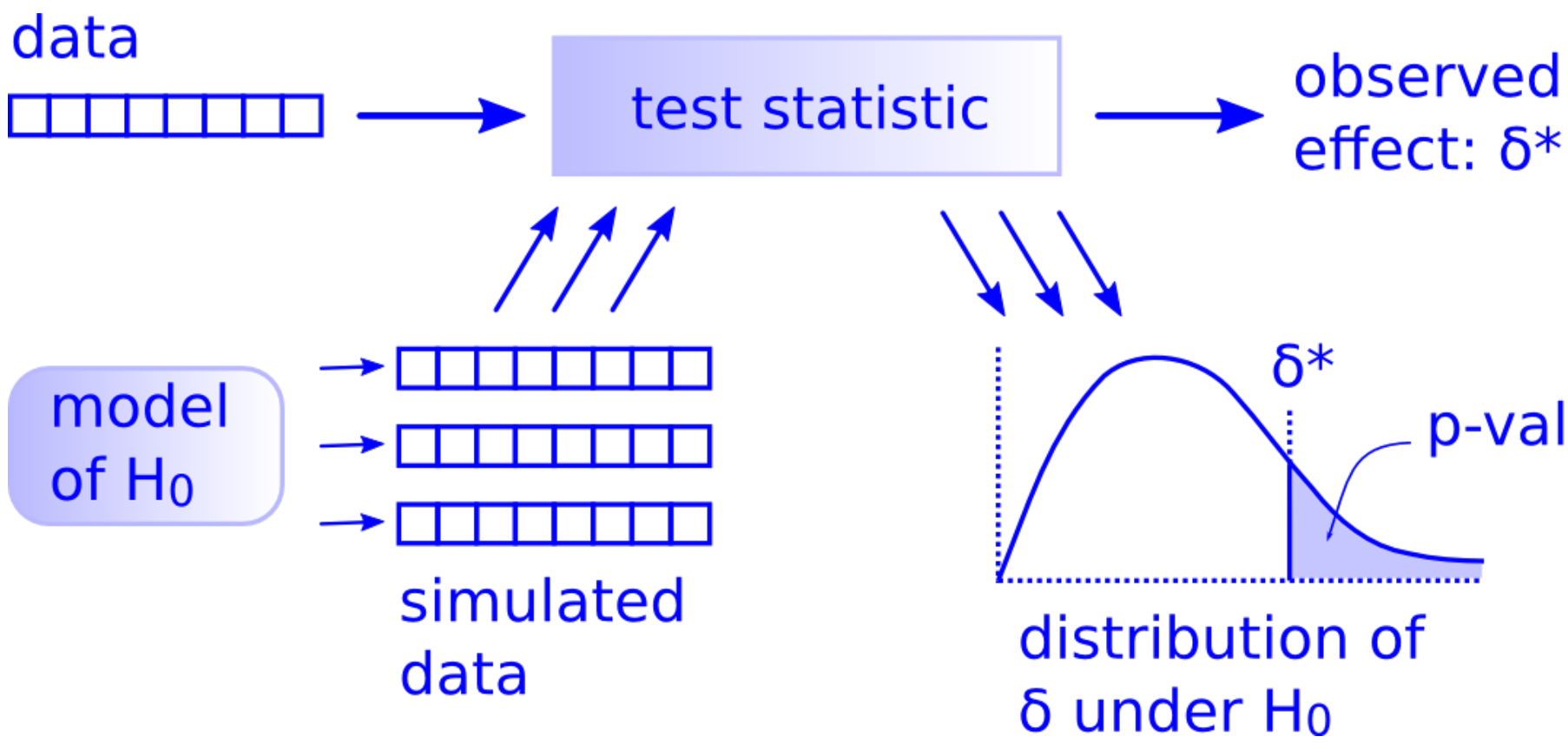
# Statistical Inference

A “you don’t need no PhD in Statistics” moment…



# Statistical Inference

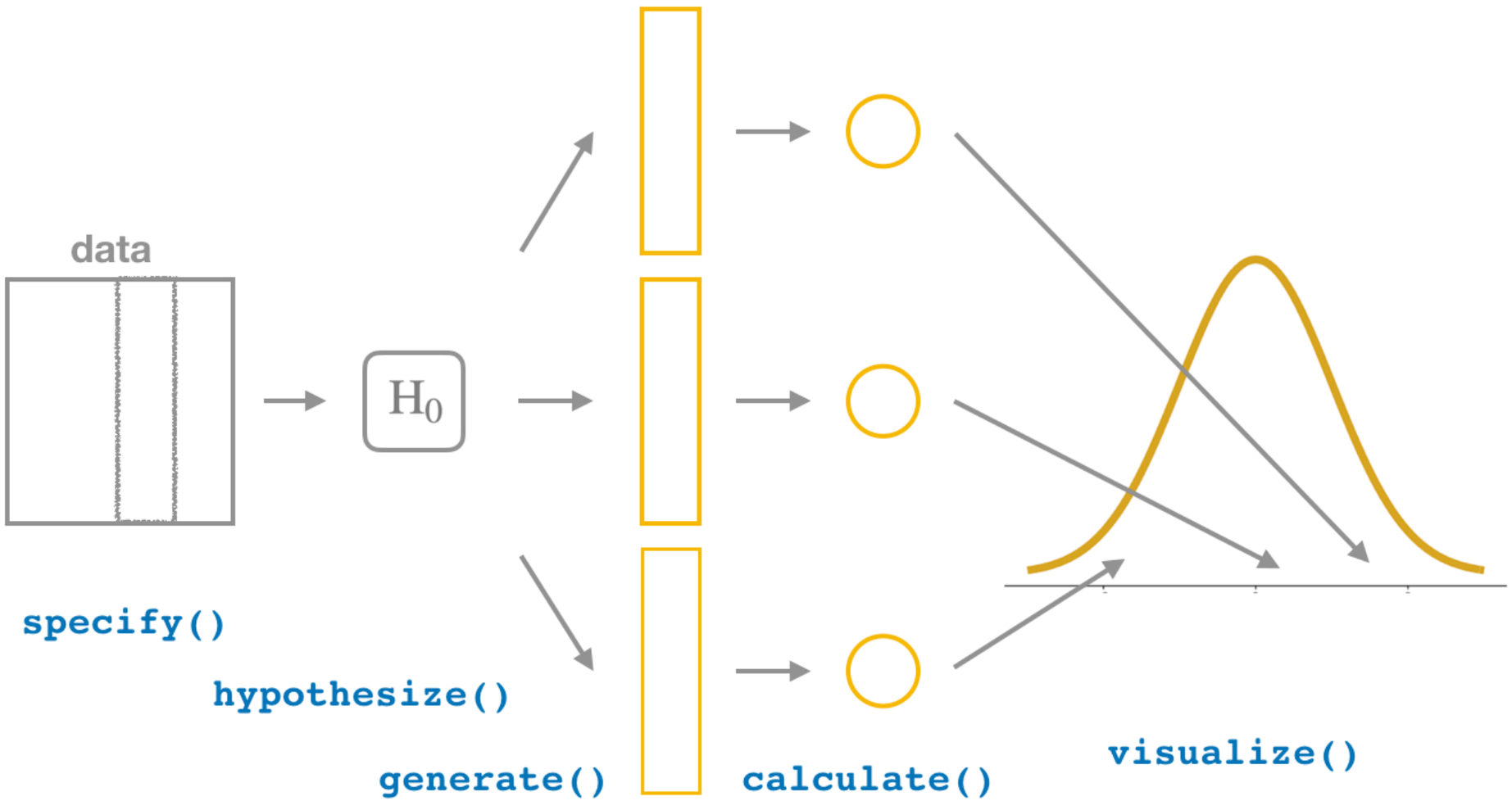
"There is only one test"




From: A.Downey blogpost

# Statistical Inference

infer package for “tidy” & transparent statistical inference

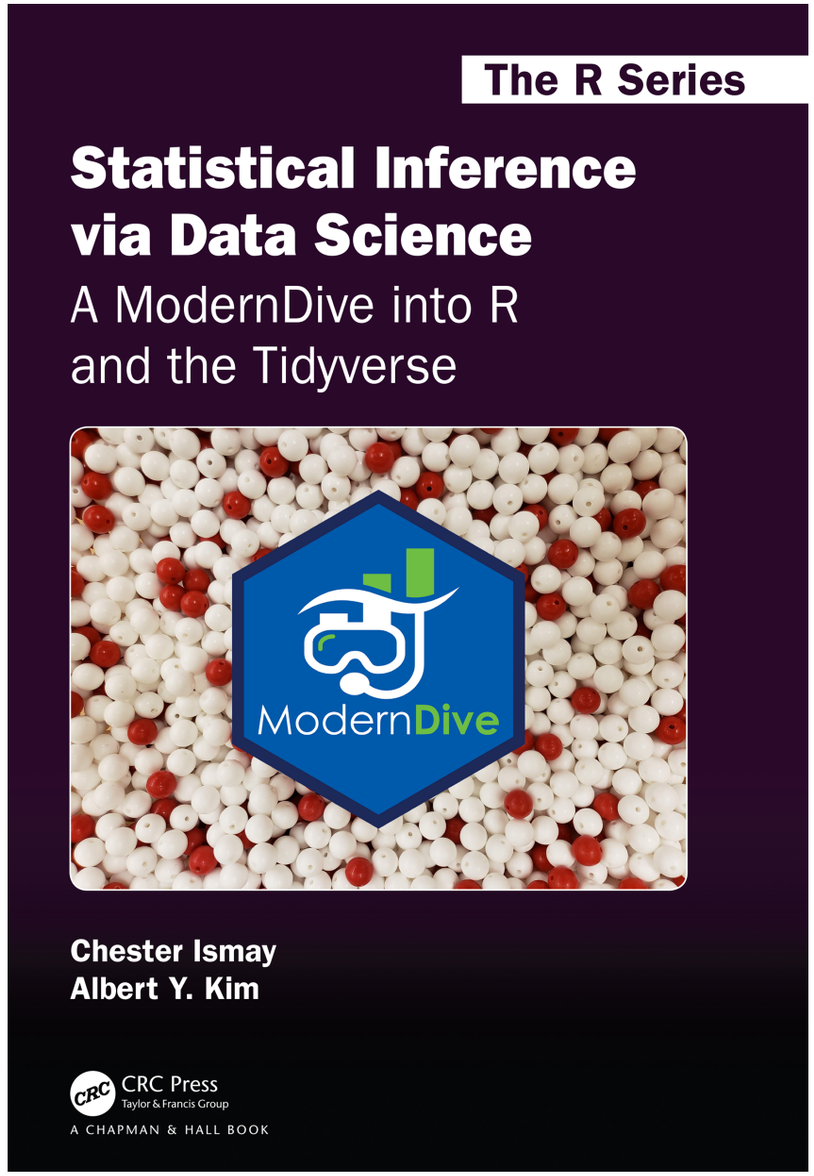



From: [Bray, Ismay, Chasnovski, Baumer, and Cetinkaya-Rundel]

# Statistical Inference

## Simulation based inference

# BTW these ideas are implemented in:




- Available online at moderndive.com
- For sale at JMM2020 Taylor & Francis booth
- CRC Press website: Use discount code ASA18