

# **Visualization Best Practices**

**Sarah McDonald**

# Overview

- Motivation: Why do we care about good graphics?
- Bad Visualization Habits
- Comments on Common Types of Charts

## What this presentation is:

- Overview of good and bad visualization practice
- Examples!

## What this presentation is not:

- ggplot tutorial
- a review of psychology of perception of visualizations etc

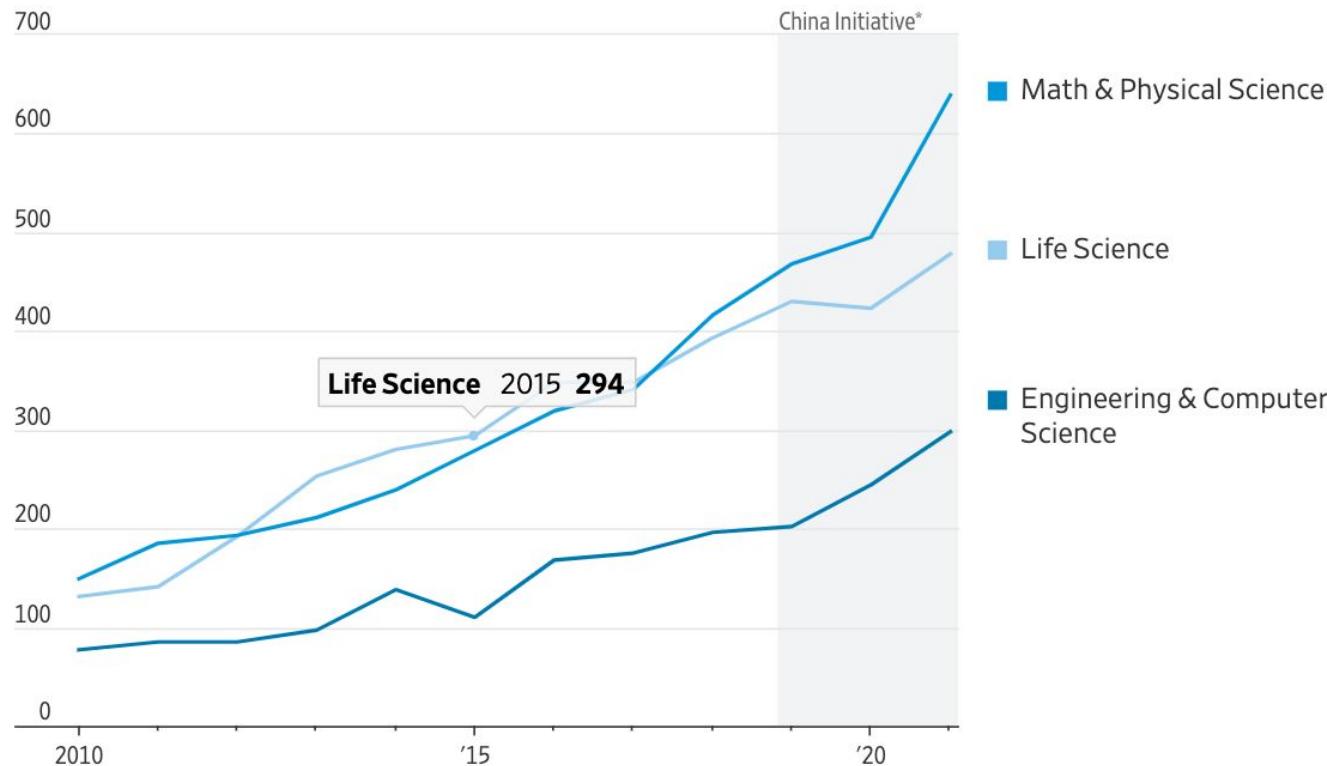
# Ethics of Visualization

# Poor visualization can

1. Misinform and underinform
2. Deceive

## Outward Bound

Number of U.S.-based scientists of Chinese descent who have left for China each year



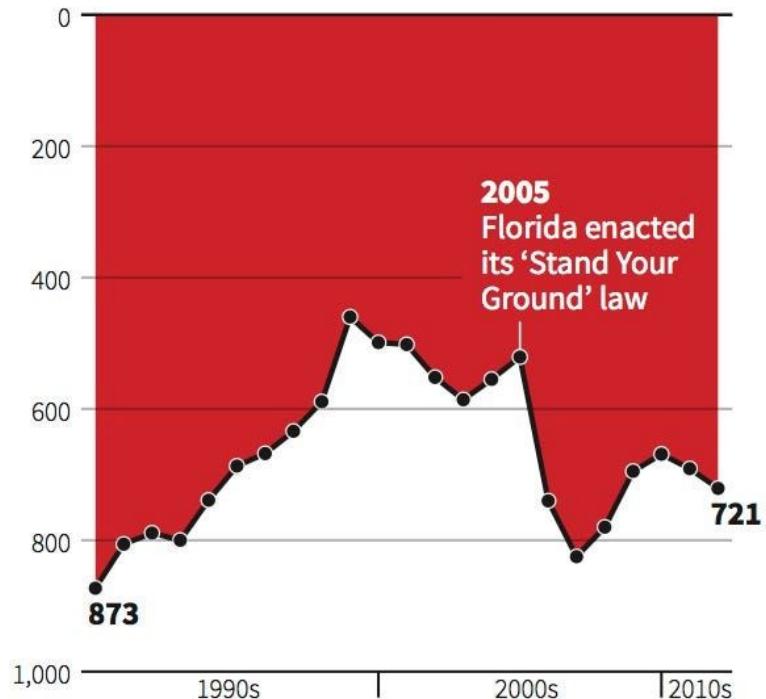
\*Note: China Initiative cases targeting academics rose in 2020.

Source: Asian American Scholar Forum

[https://www.wsj.com/articles/u-s-china-tensions-fuel-outflow-of-chinese-scientists-from-u-s-universities-11663866938?st=ydbriz4x66ive54&reflink=desktopwebshare\\_permalink](https://www.wsj.com/articles/u-s-china-tensions-fuel-outflow-of-chinese-scientists-from-u-s-universities-11663866938?st=ydbriz4x66ive54&reflink=desktopwebshare_permalink)

# Gun deaths in Florida

Number of murders committed using firearms



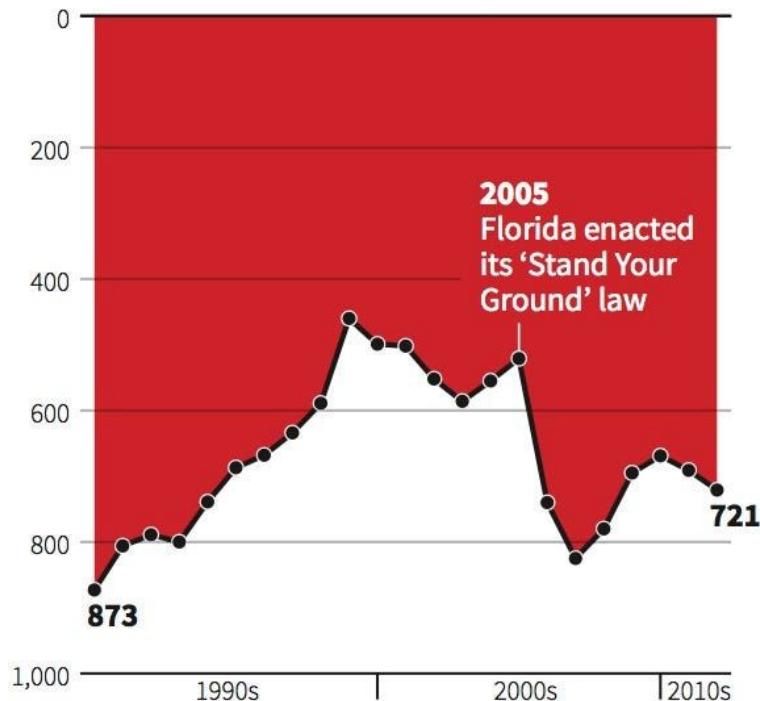
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

# Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014



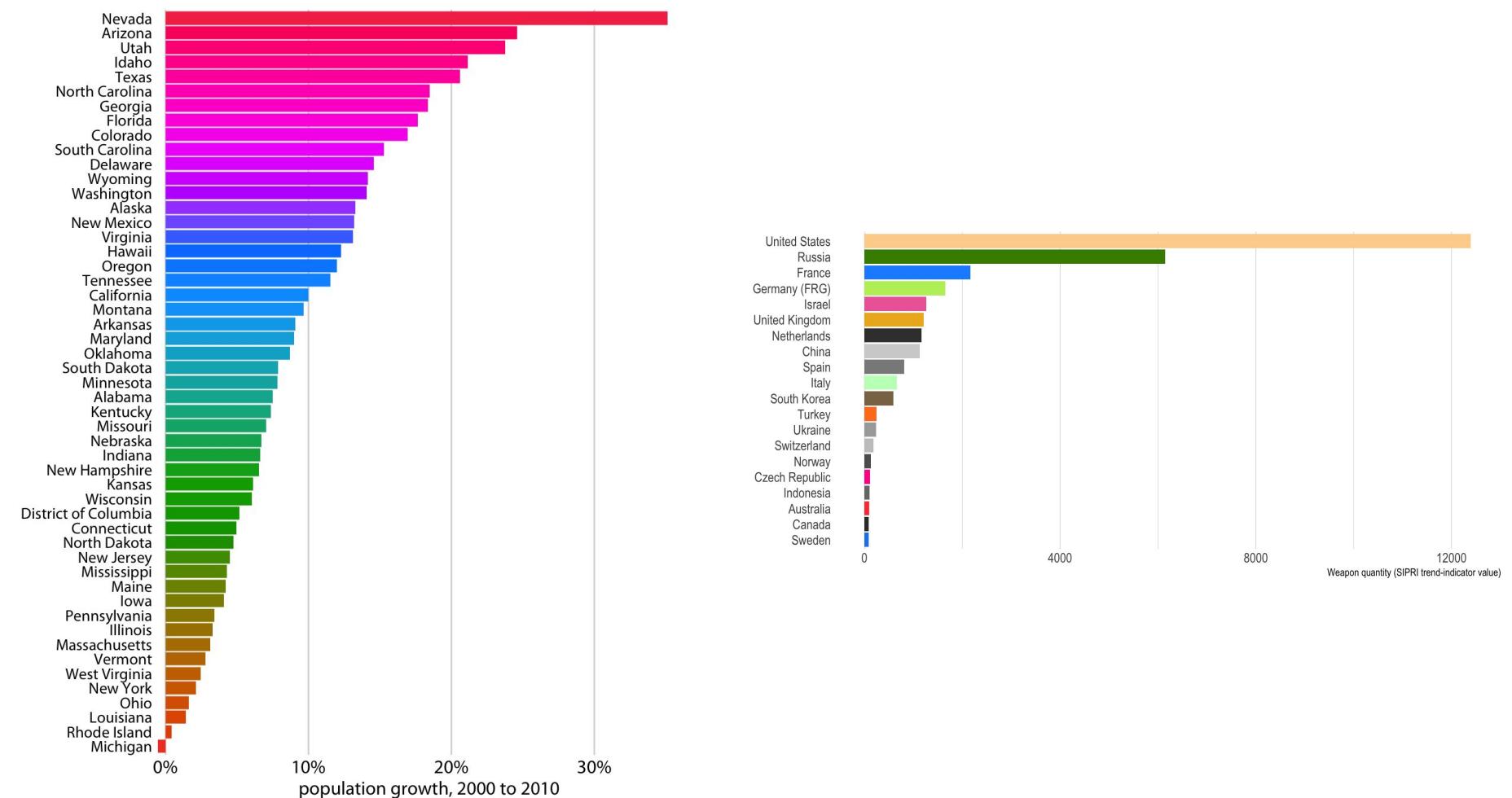
## Gun deaths in Florida

# Some bad visualization habits

# Colors

# Colors

- Use colors to serve a purpose (distinguishing groups, embellishment, scales)
- Use continuous colors for continuous variables, discrete colors for discrete variables
- Try to use colorblind-friendly colors

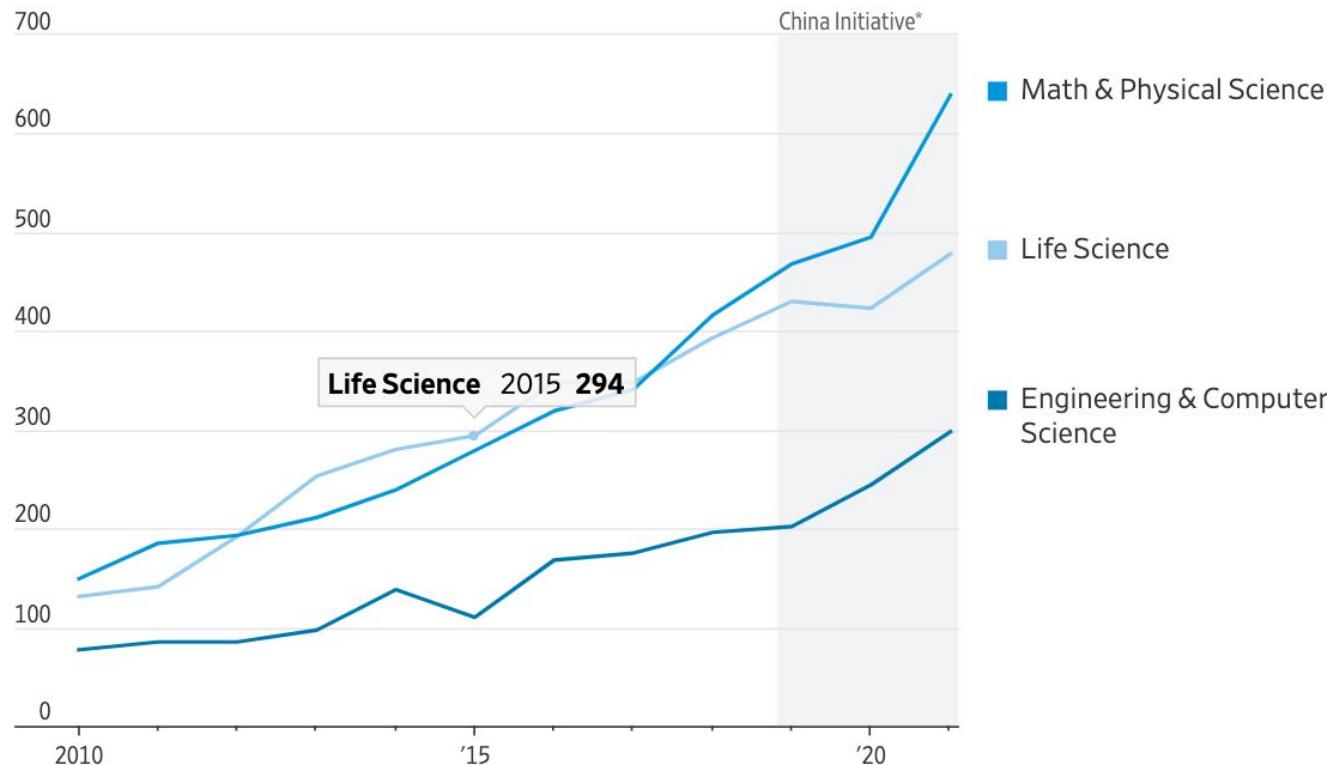


<https://clauswilke.com/dataviz/color-pitfalls.html#using-non-monotonic-color-scales-to-encode-data-values>

[https://www.data-to-viz.com/caveat/color\\_com\\_nothing.html](https://www.data-to-viz.com/caveat/color_com_nothing.html)

## Outward Bound

Number of U.S.-based scientists of Chinese descent who have left for China each year

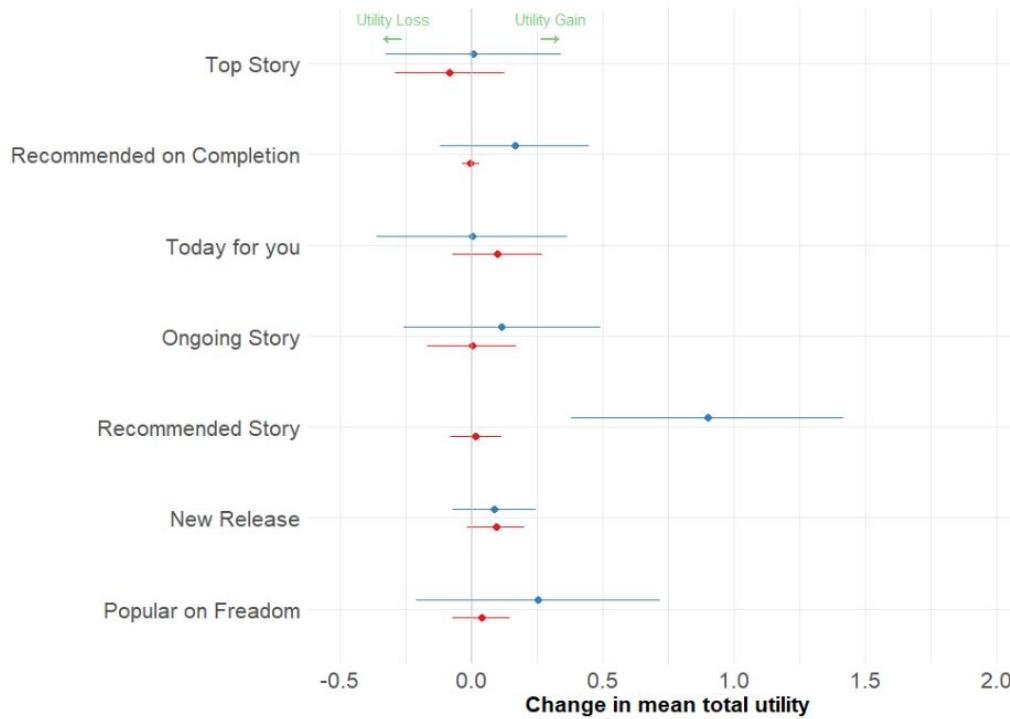


\*Note: China Initiative cases targeting academics rose in 2020.

Source: Asian American Scholar Forum

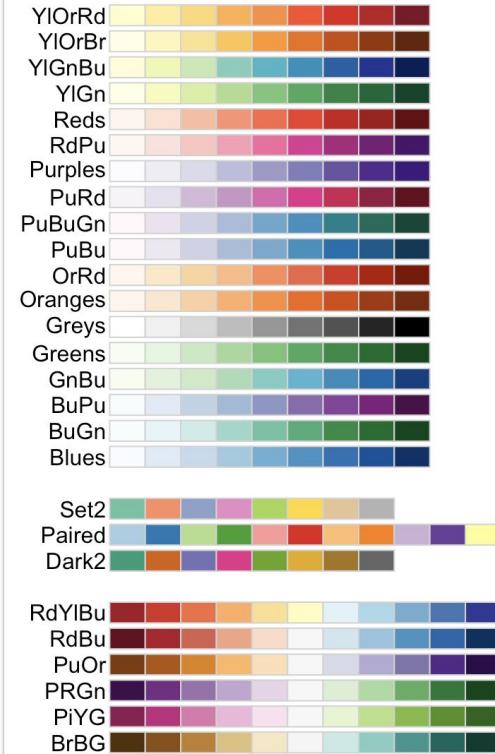
[https://www.wsj.com/articles/u-s-china-tensions-fuel-outflow-of-chinese-scientists-from-u-s-universities-11663866938?st=ydbriz4x66ive54&reflink=desktopwebshare\\_permalink](https://www.wsj.com/articles/u-s-china-tensions-fuel-outflow-of-chinese-scientists-from-u-s-universities-11663866938?st=ydbriz4x66ive54&reflink=desktopwebshare_permalink)

**Figure 4:** Difference in average total utility in treatment and control groups for eight most popular trays.

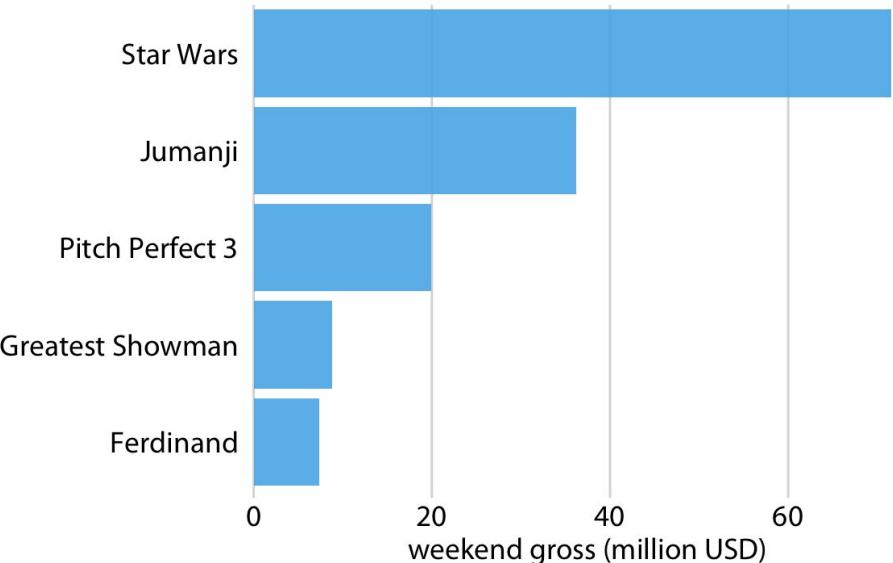
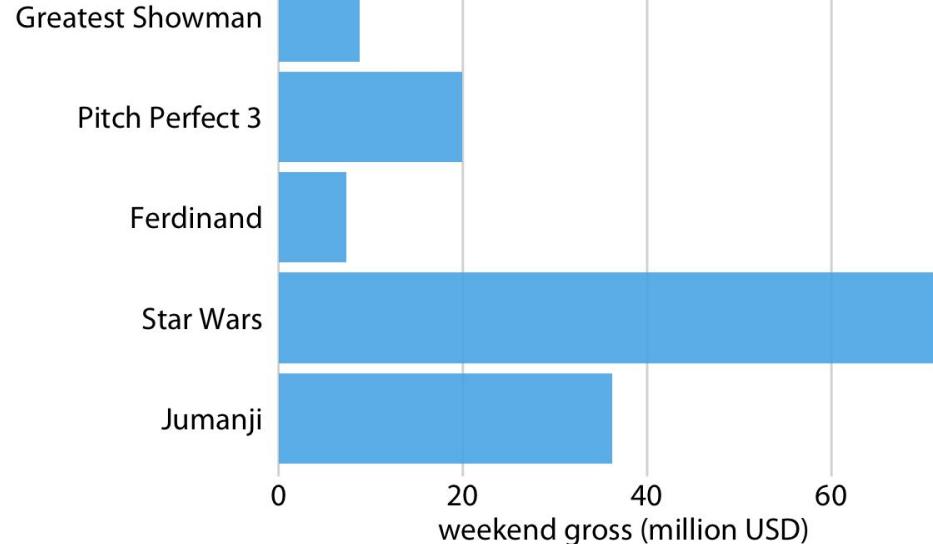


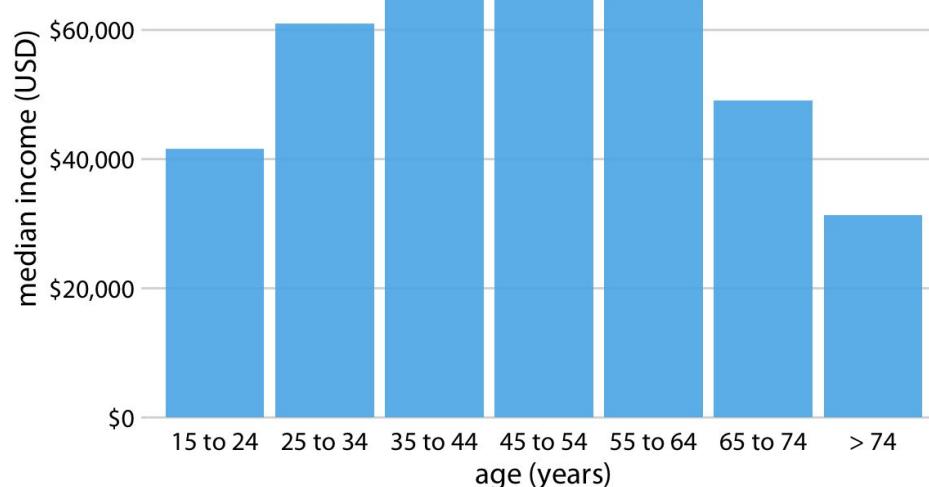
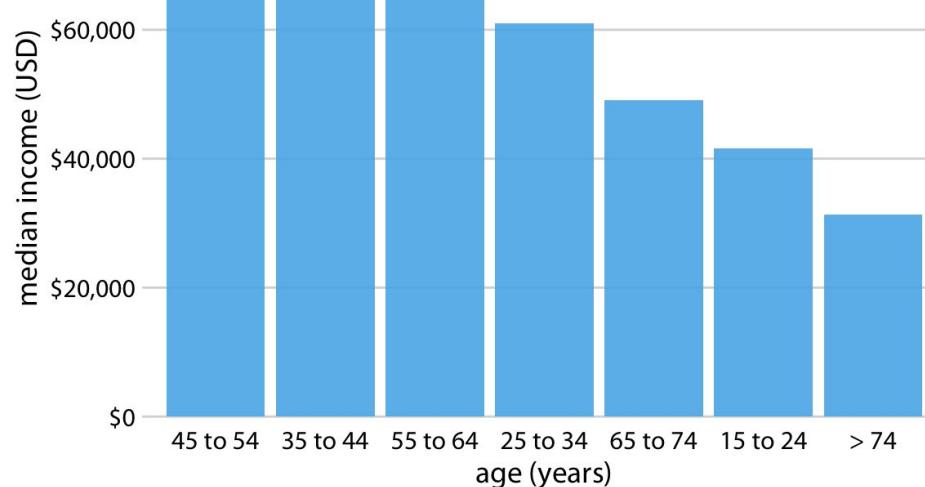
*Note: Difference in average total utility in treatment and control groups for eight most popular trays. Experimental period in blue, pre-experimental in red. Pre-experimental period is 7-19.06.2021, there are approximately twice as many users in the per-experimental period (this date is chosen on the basis of being the closest two-weeks long period without other major experiments and alterations in the app).*

```
display.brewer.all(colorblindFriendly = TRUE)
```



# Ordering





# Axes/Scales/Text

# Axes/Scales

## Scales

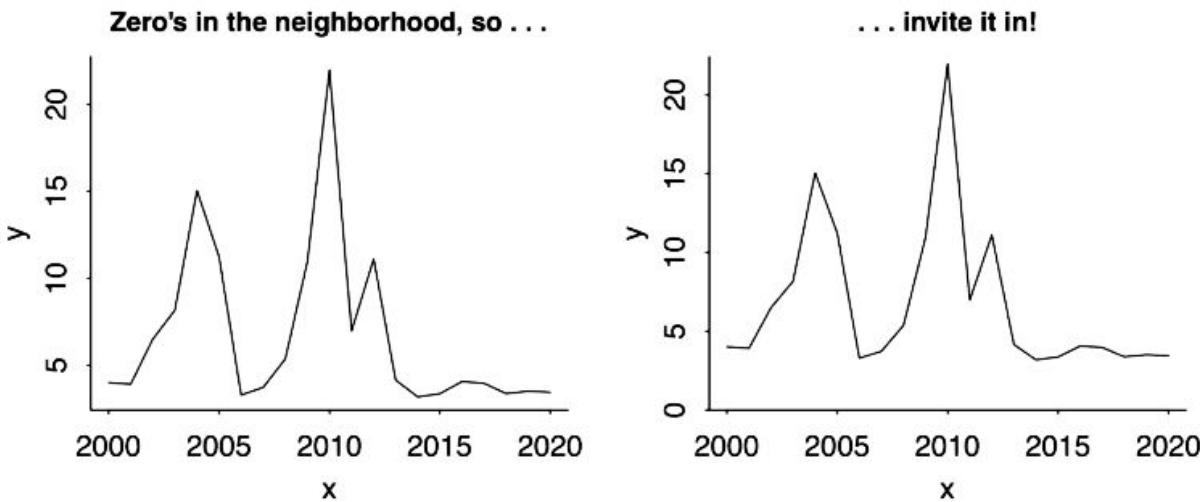
- "If zero is in the neighborhood, invite it in."
- Evenly spaced multiples of 1s, 2s, or 5s
- Don't truncate bar chart axes
- Don't use double axes

## Axis Labels

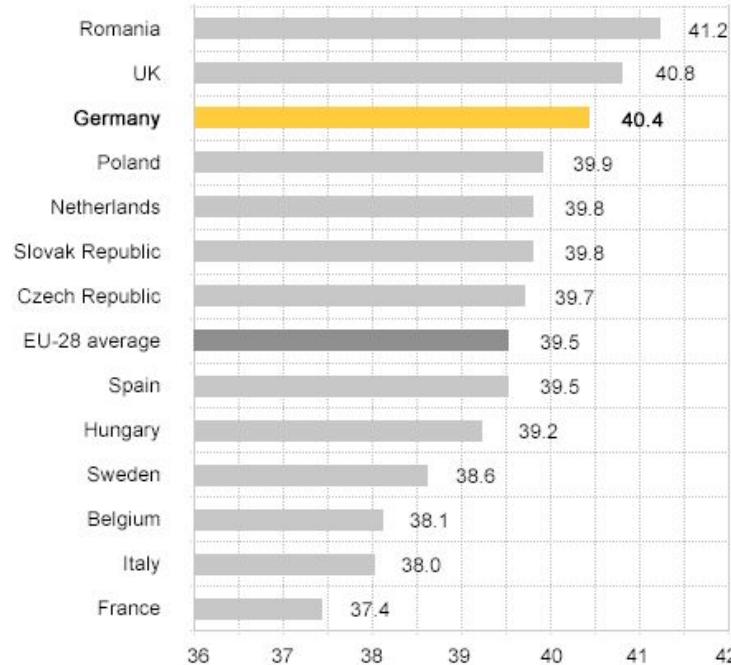
- Don't make people turn their head
- Don't overcrowd

## Axis Titles

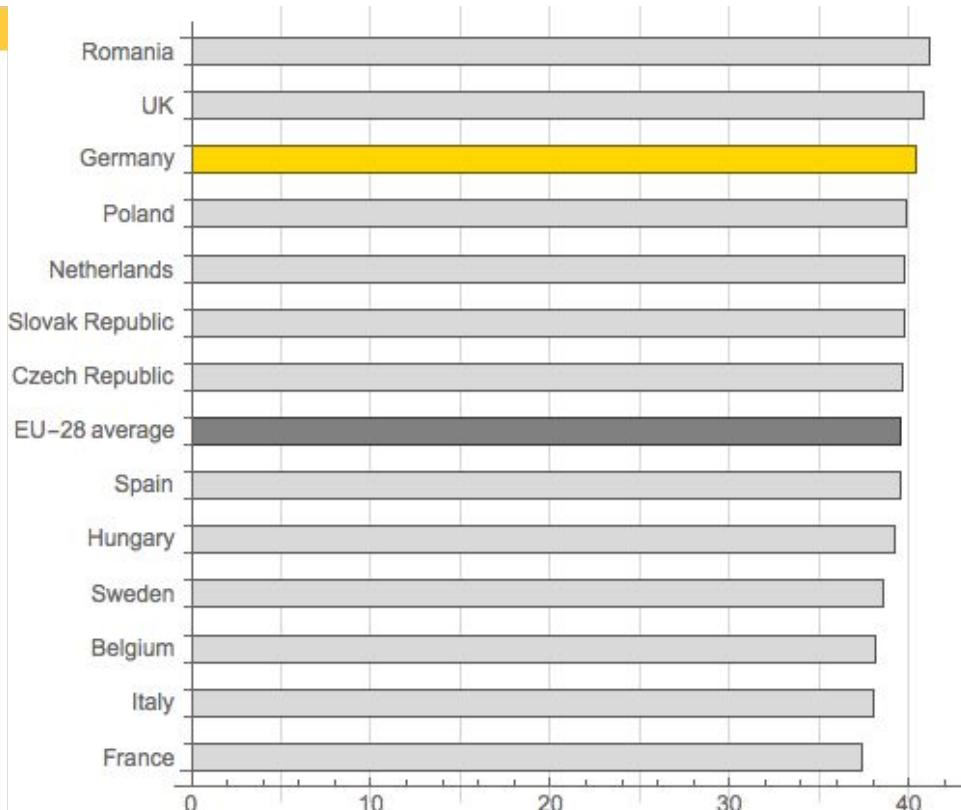
- Maybe rotate your y-axis title or remove it (a little controversial)



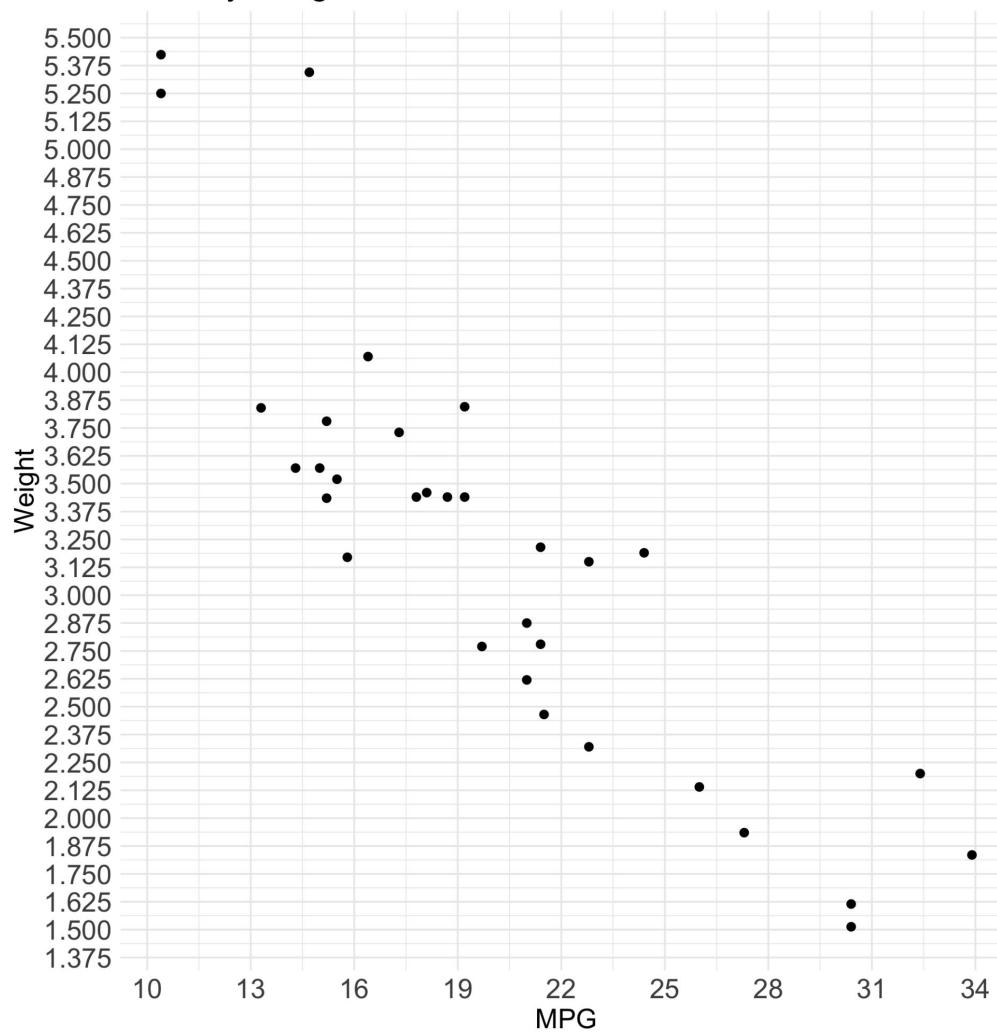
Average number of actual weekly hours of work in main job, full-time employees, 2013

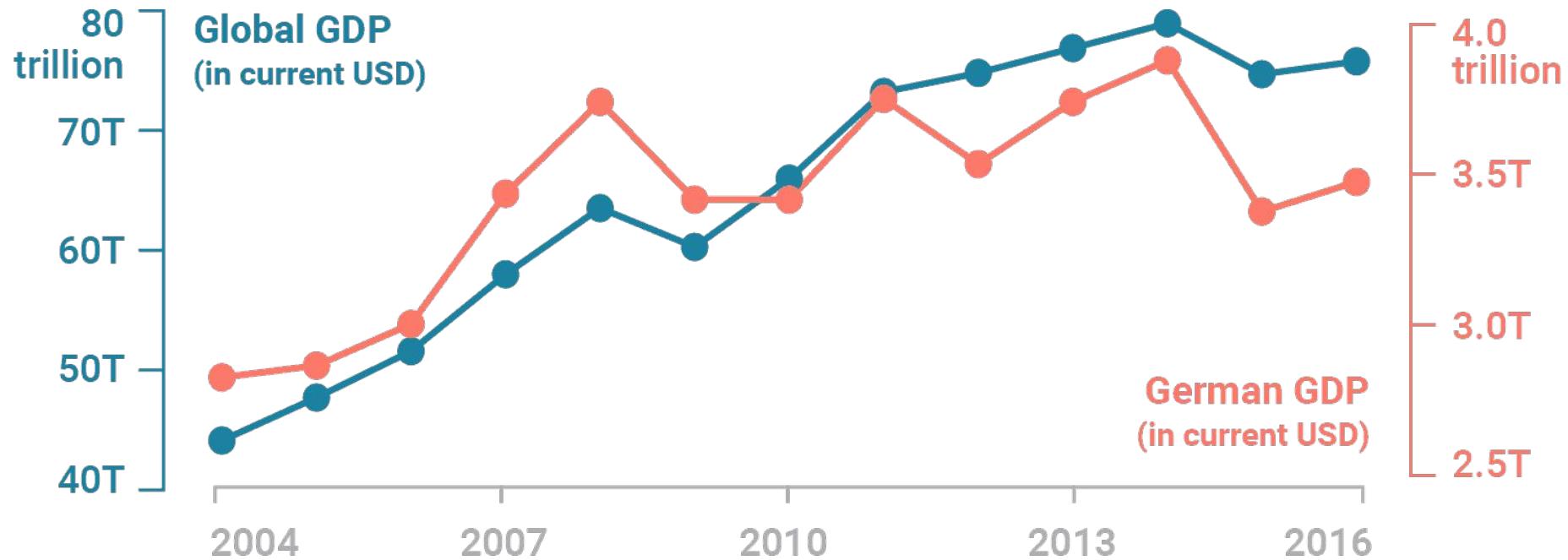


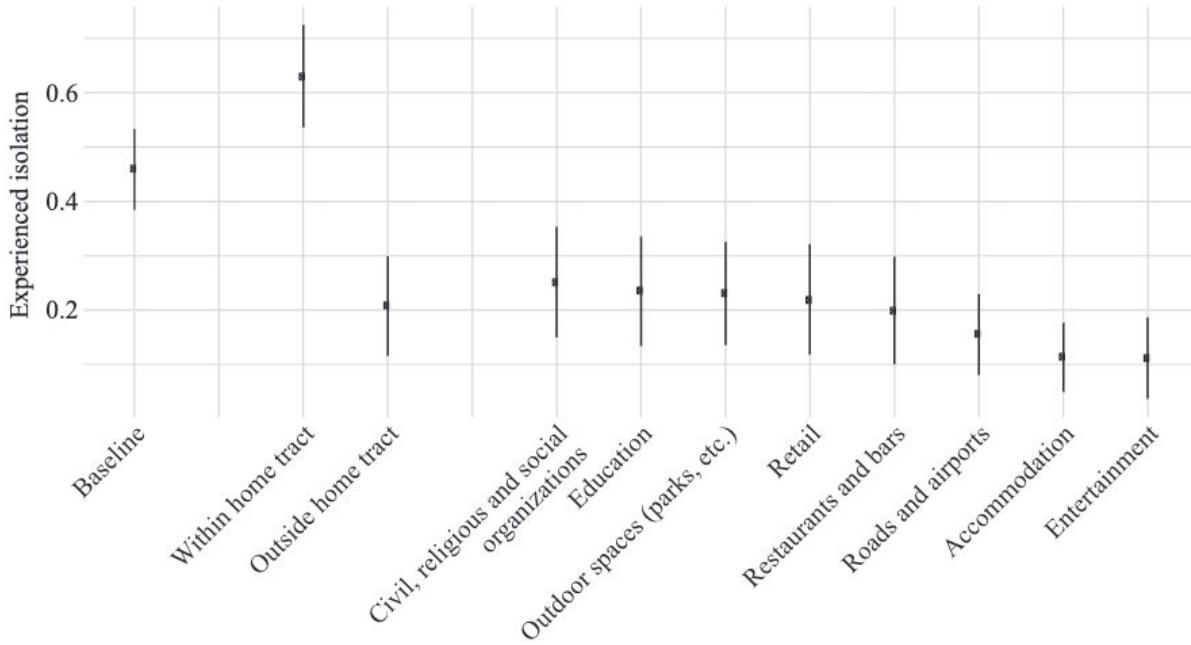
Source: Eurofound 2014



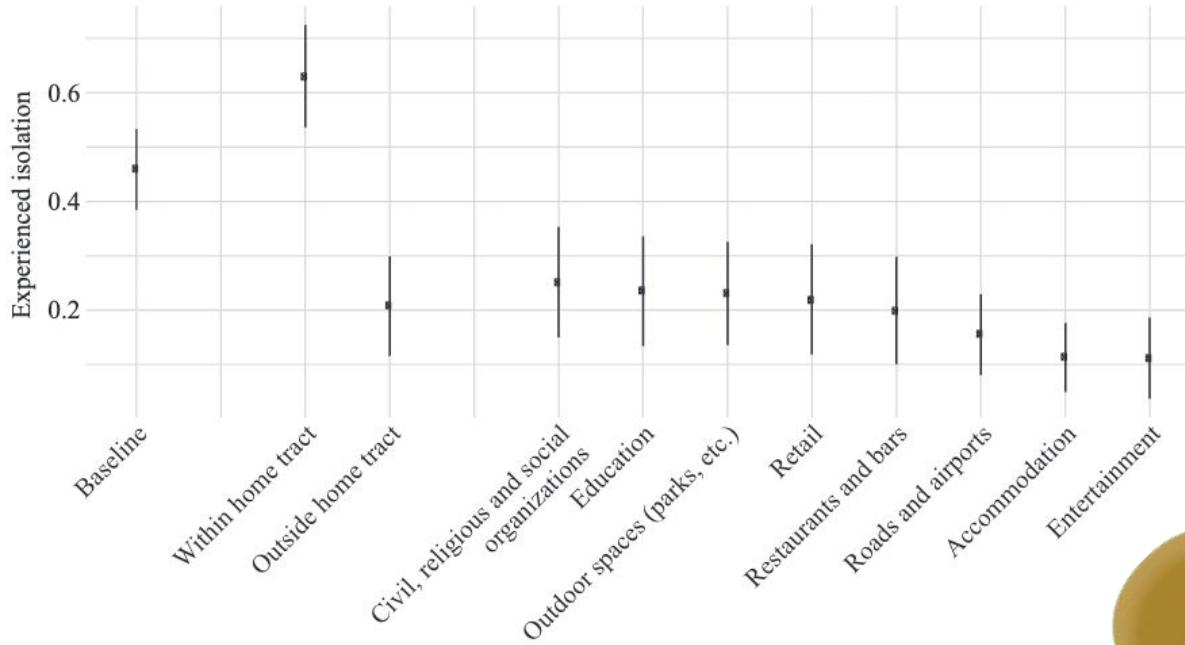
# MPG by Weight







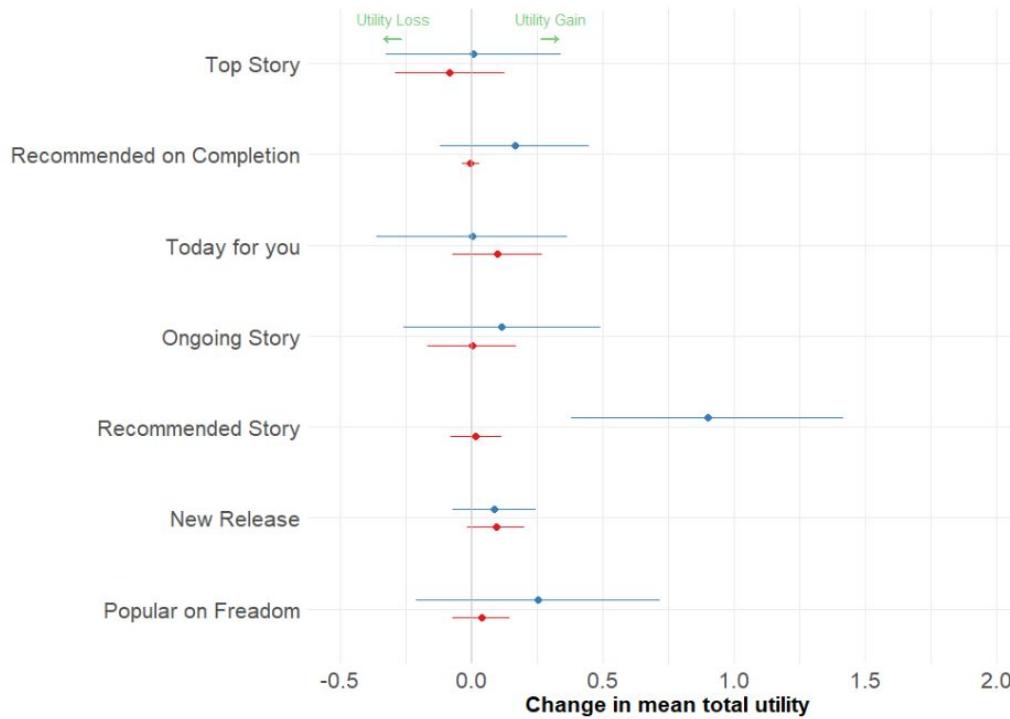
**Fig. 5.** Experienced isolation relative to baseline by location. We plot the population-weighted mean experienced isolation in a particular feature and compare it with our baseline measure. Error bars show the population-weighted SD of experienced isolation across MSAs.



**Fig. 5.** Experienced isolation relative to baseline by location. We plot the population-weighted mean experienced isolation for each location and compare it with our baseline measure. Error bars show the population-weighted SD of experienced isolation across MSAs.

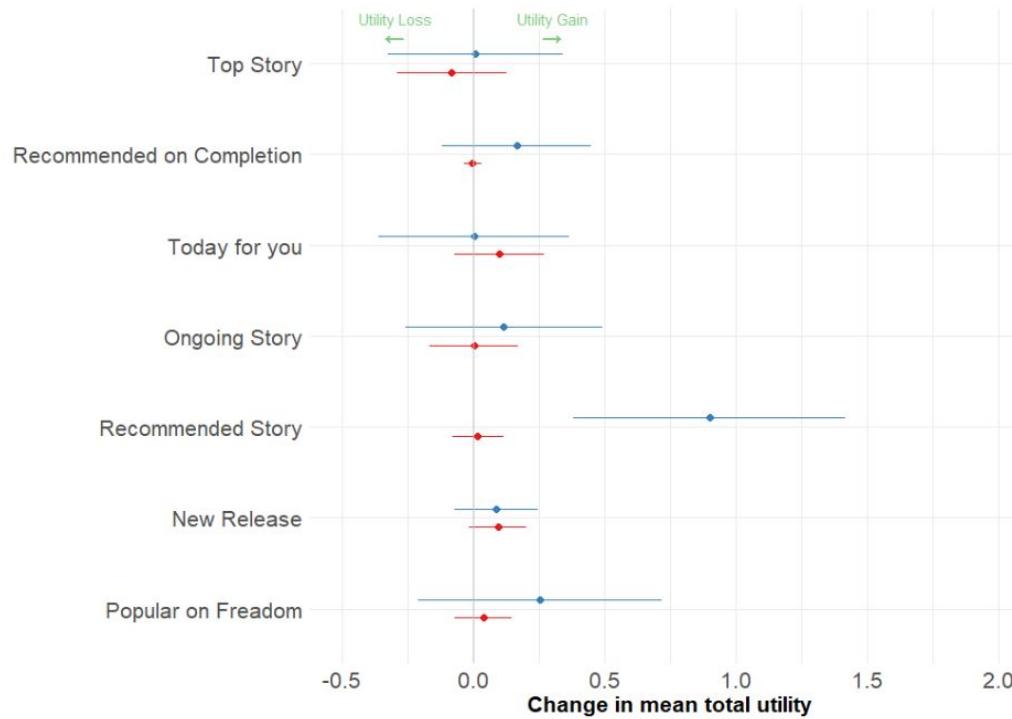
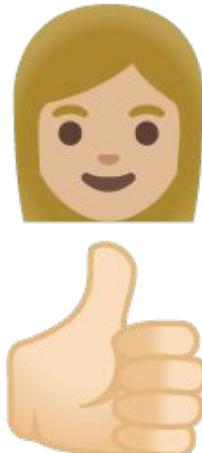


**Figure 4:** Difference in average total utility in treatment and control groups for eight most popular trays.

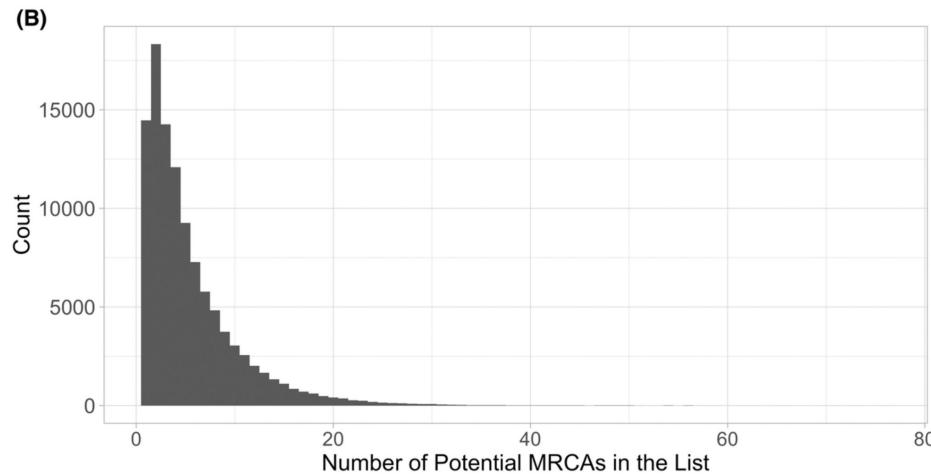
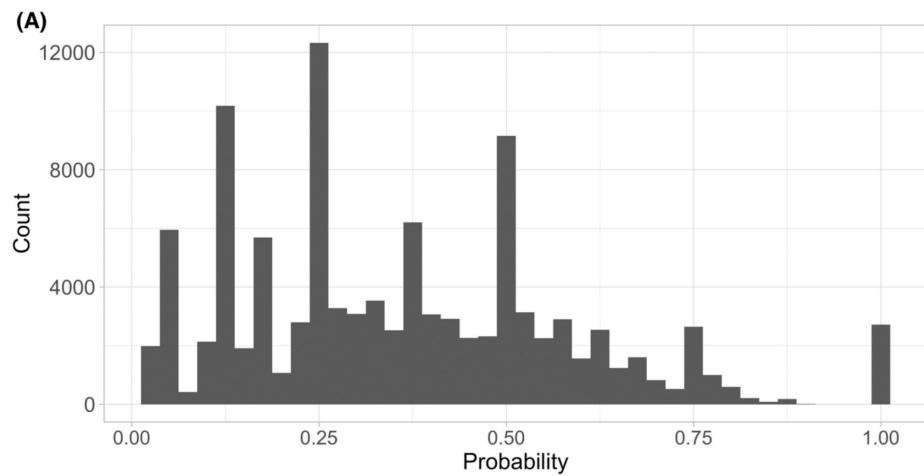


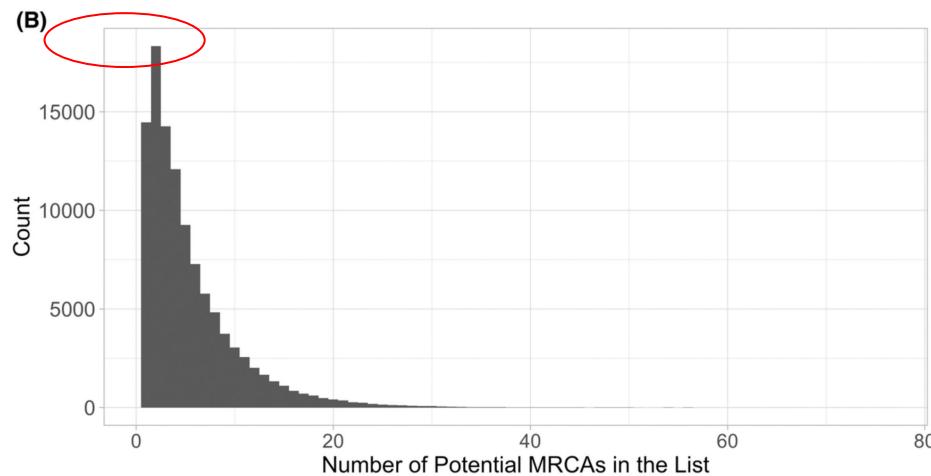
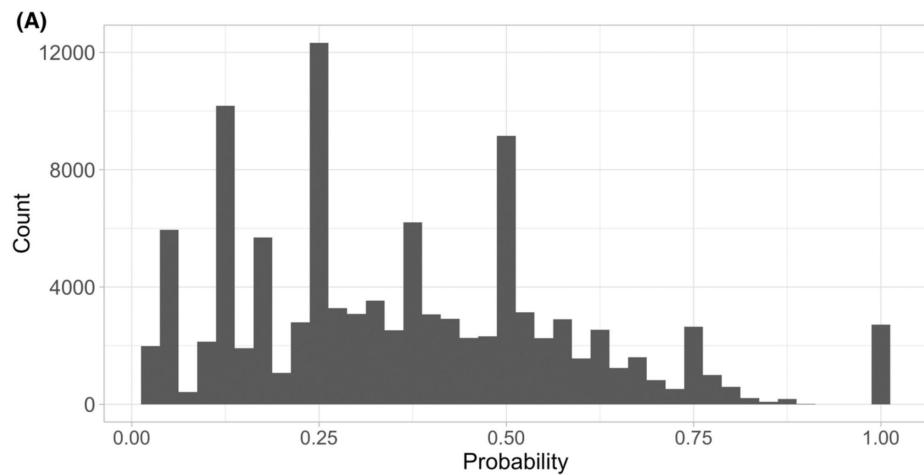
*Note: Difference in average total utility in treatment and control groups for eight most popular trays. Experimental period in blue, pre-experimental in red. Pre-experimental period is 7-19.06.2021, there are approximately twice as many users in the per-experimental period (this date is chosen on the basis of being the closest two-weeks long period without other major experiments and alterations in the app).*

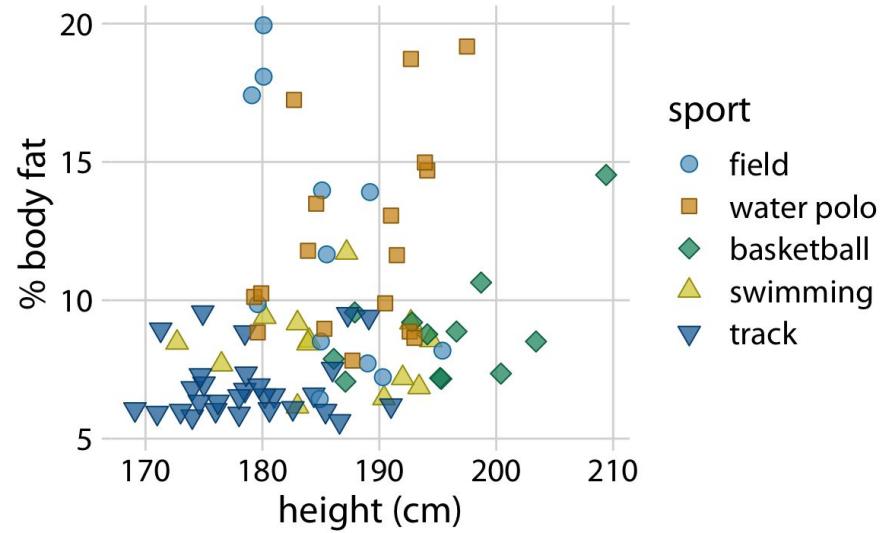
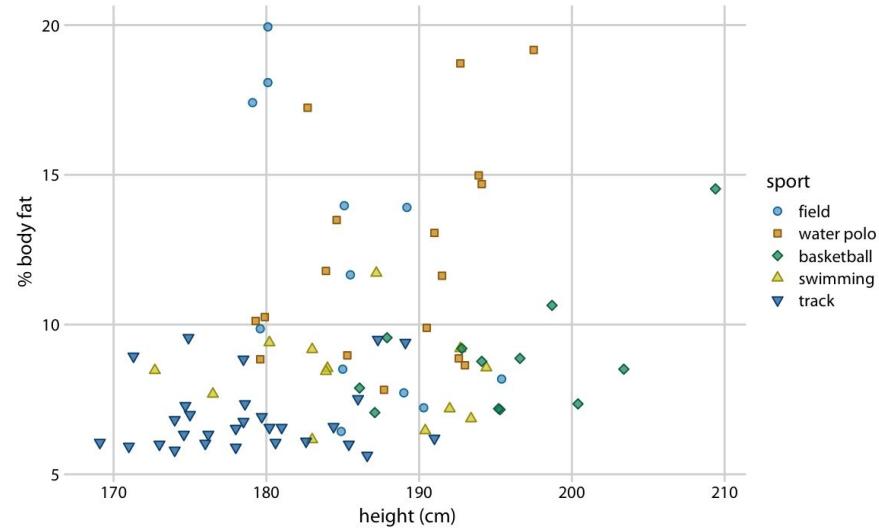
**Figure 4:** Difference in average total utility in treatment and control groups for eight most popular trays.



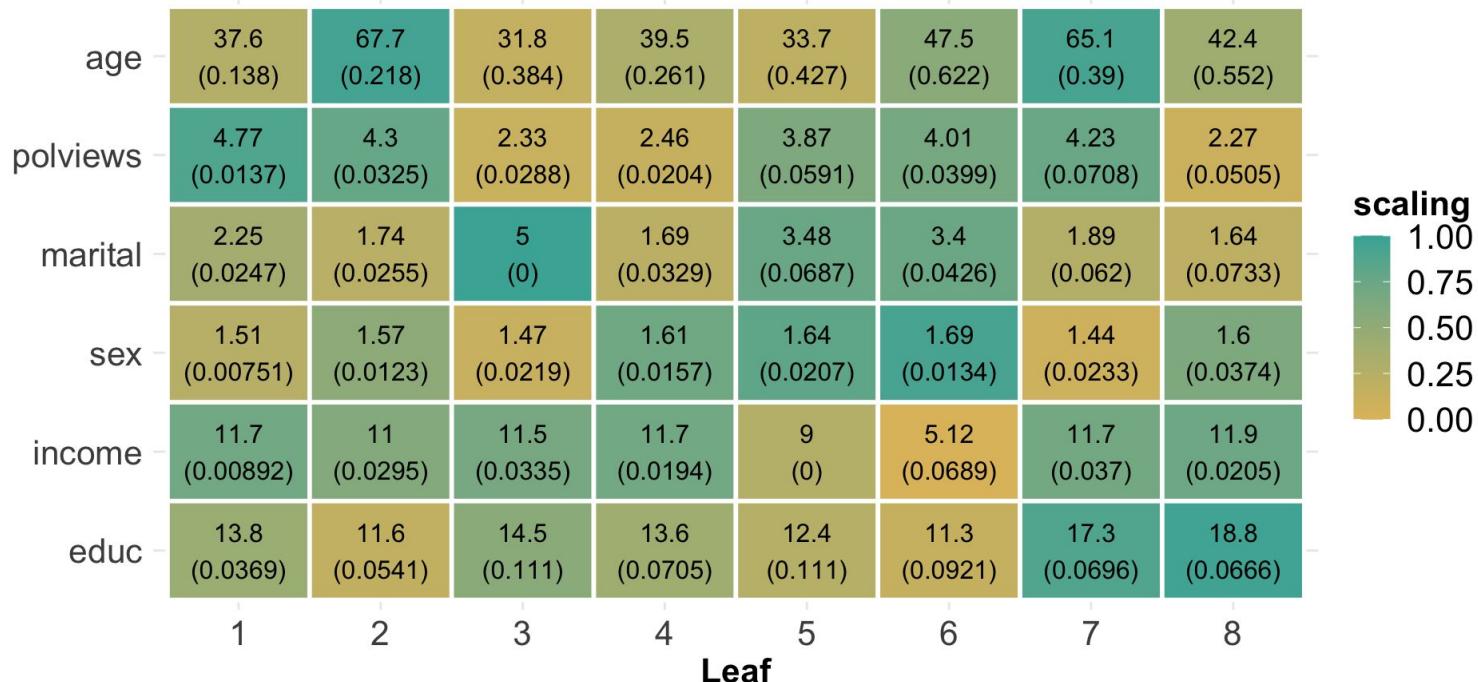
*Note: Difference in average total utility in treatment and control groups for eight most popular trays. Experimental period in blue, pre-experimental in red. Pre-experimental period is 7-19.06.2021, there are approximately twice as many users in the per-experimental period (this date is chosen on the basis of being the closest two-weeks long period without other major experiments and alterations in the app).*



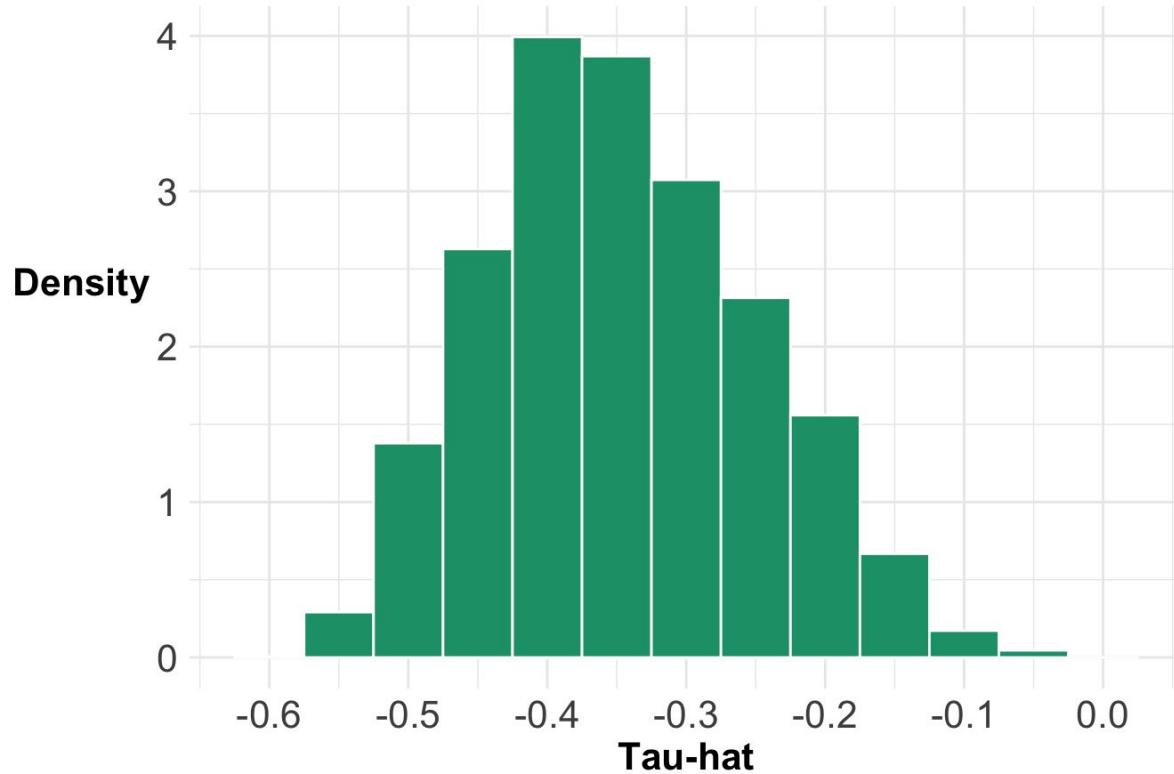




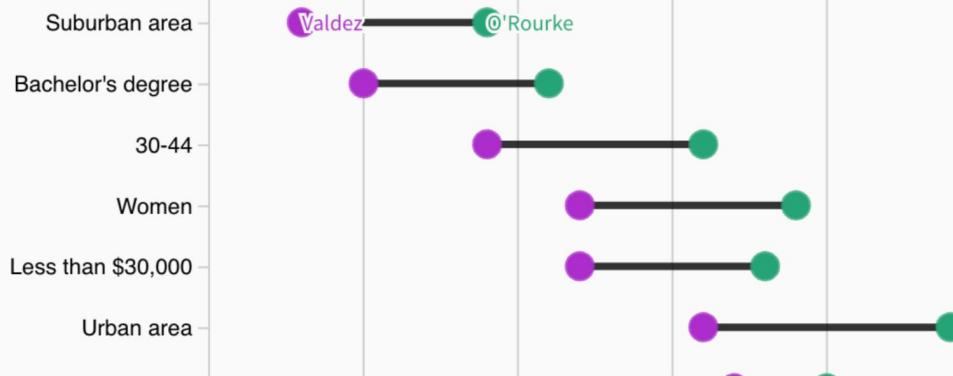
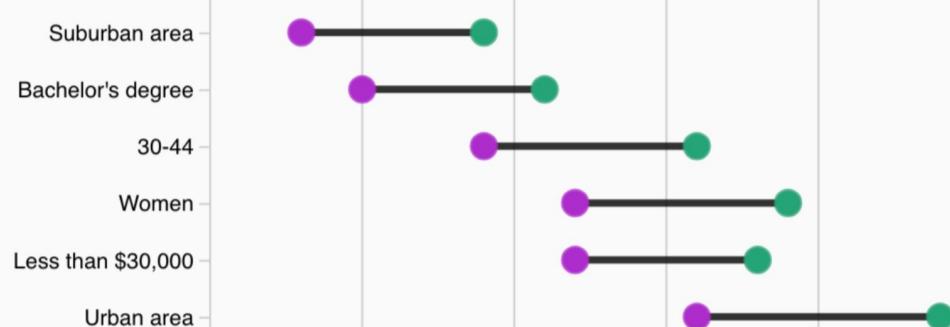
## Average covariate values within each leaf



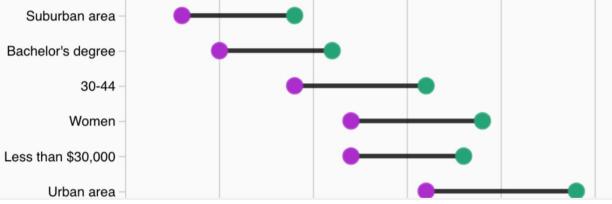
## CATE Estimates



O'Rourke Valdez

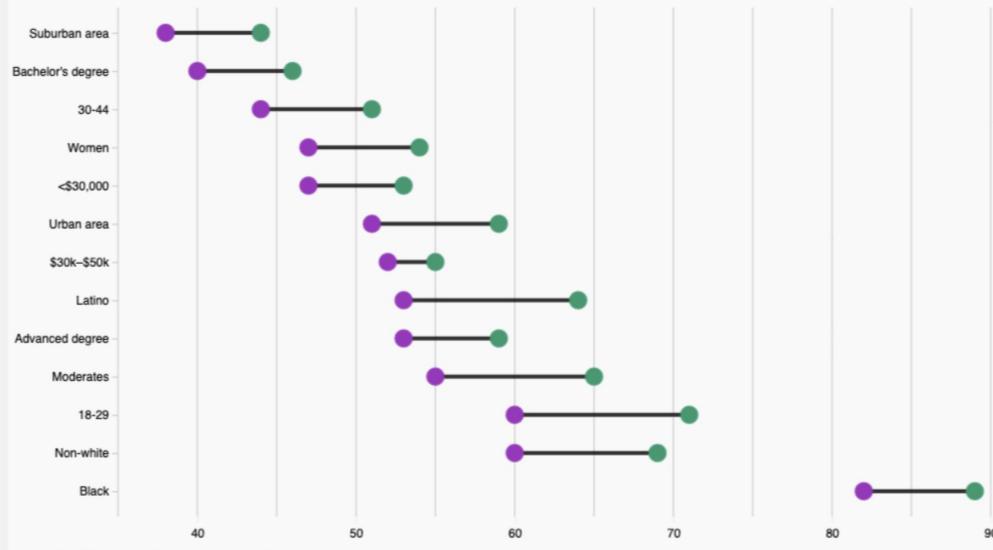


O'Rourke Valdez



### In Texas, Beto O'Rourke outperformed Lupe Valdez among key Democratic groups

Votes margin in 2018 election



Source: CNN/Edison exit polls + Gubernatorial and Senate races

# Context

# Context

- If your paper is in overleaf, the text in your figures should match the overleaf font
  - Use package [extrafont](#)
- Thematic elements (colors, fonts) should be the same throughout entire paper or presentation
  - Make your own theme function and add it to every plot
    - Don't use default settings/colors
    - Predefine the colors you use for the entire script
- Use the same plotting packages/software for all figures
  - Don't switch between base R and ggplot2

Panel B. Contribution to net imports

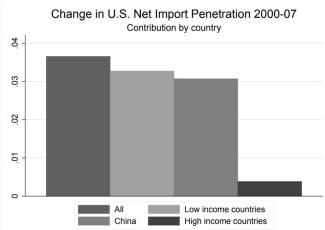


Figure 3

**Note:** This figure presents the change in U.S. import penetration (panel A) and net import penetration (panel B) from 2000 to 2007. Import penetration is measured as the ratio of imports to U.S. expenditures themselves measured as domestic shipments plus net imports. We decompose the change in import penetration by countries: low income countries (including China), China, and high income countries.

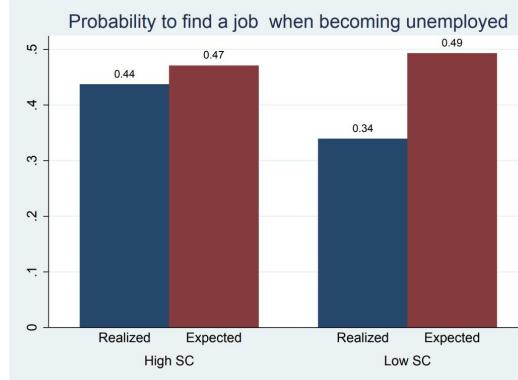


Figure 7

#### Realized and Expected Duration of Unemployment Spells

**Note:** This figure presents realized and expected duration of unemployment spells. We draw from the Health and Retirement Study (HRS), a longitudinal survey conducted every two years. Individuals are asked about their current job status (employed, unemployed, retired), and about their expectations of future labor outcome. In particular, they are asked: "Suppose you were to lose your job this month. What do you think are the chances that you could find an equally good job in the same line of work within the next few months?" Red bars present the average perceived probability to find a job after becoming unemployed, computed across participants in the HRS waves of 2000, 2002 and 2004. Blue bars present the probability that an individual who was employed in year 2000 (according to HRS), but not in year 2002, finds a job in year 2004. High SC (respectively Low SC) denote commuting zones that lie in the top tercile (respectively bottom tercile) of the distribution of shipping costs. Information from the County Business Patterns datasets on the structure of employment across 6-digit NAICS industries by CZs is used to compute Shipping costs at the CZ level.

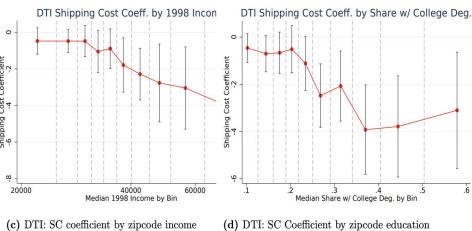


Figure 6

#### Heterogeneous Treatment Effects

**Note:** This figure presents the point estimates and confidence intervals of cross-sectional regressions of the change in the debt-to-income ratio from 2000Q4 to 2007Q4 on shipping costs, our proxy for import competition, at the individual level. The specifications are similar to column (10) of Table 6 and are run separately across deciles of individual age (a), individual credit score (b), zip code income (c), and zip code share of the population with at least college education (d).

USA Trade Exposure  
Average shipping costs by Commuting Zones

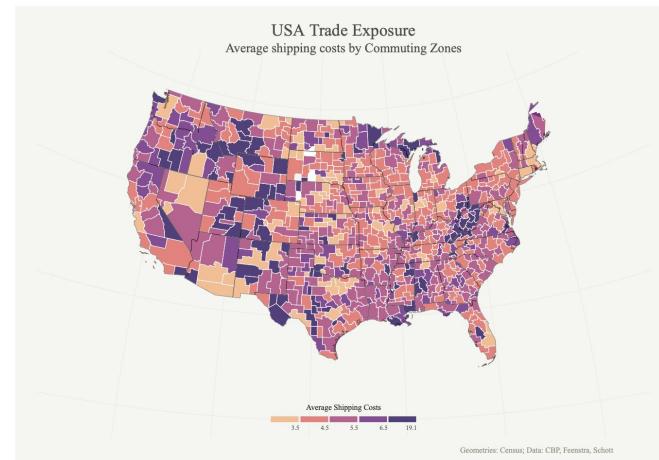


Figure 5

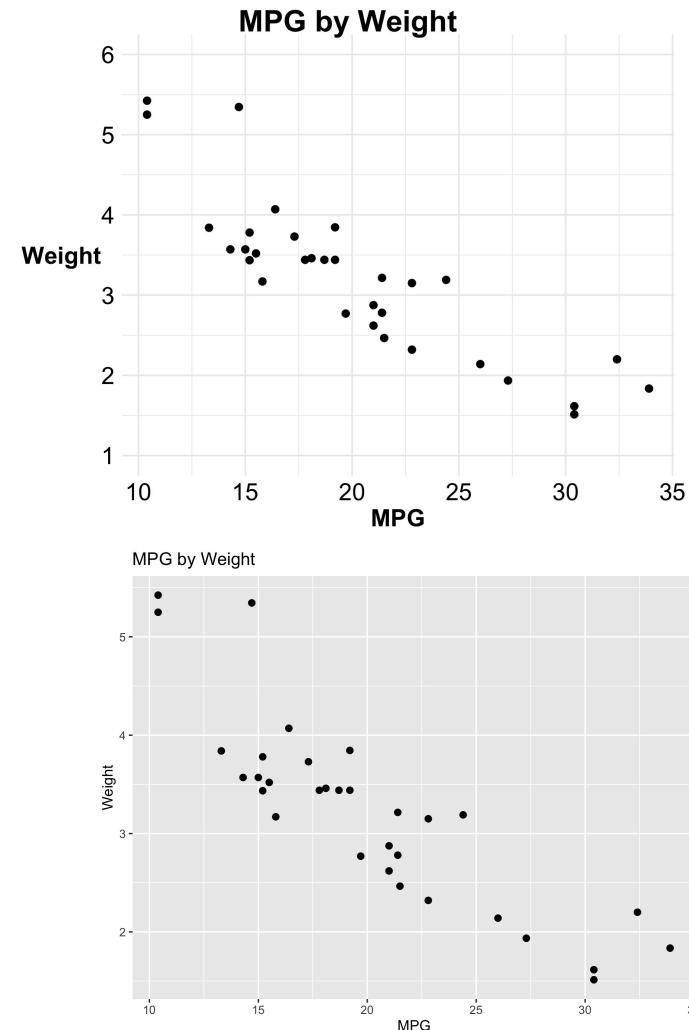
#### Average Shipping Costs by Commuting Zones

**Note:** This figure presents the distribution of shipping costs (%) across commuting zones measured in 1998. Information from the County Business Patterns datasets on the structure of employment across 6-digit NAICS industries by CZs is used to compute Shipping costs at the CZ level.

```

theme_sarah <- function(){
  # theme_minimal makes background of plot white
  # and does not distract from the figure
  theme_minimal() %+replace%
  theme(
    # set facet label size
    strip.text = element_text(size = 16),
    # set axis text sizes
    axis.text=element_text(size=16),
    # set axis text sizes and set to bold
    axis.title=element_text(size=16,face="bold"),
    axis.title.y = element_text(angle = 0, vjust = 0.5),
    # increase space between facets
    panel.spacing = unit(1, "lines"),
    # set legend title sizes and set to bold
    legend.title = element_text(size=16,face="bold"),
    # set legend text sizes
    legend.text = element_text(size=16),
    # set plot text sizes
    plot.title = element_text(size = 20, face = "bold"),
    # set plot title position to left corner
    plot.title.position = "plot"
  )
}

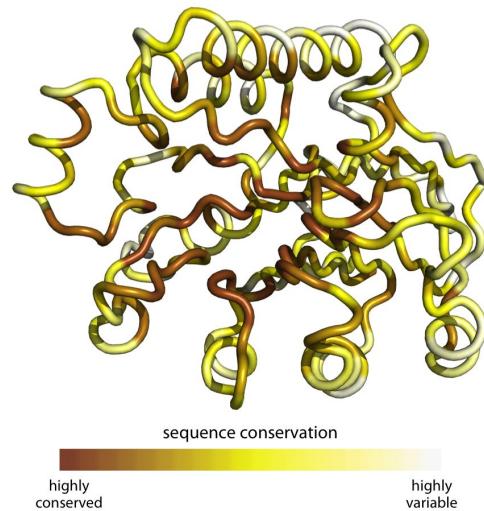
```



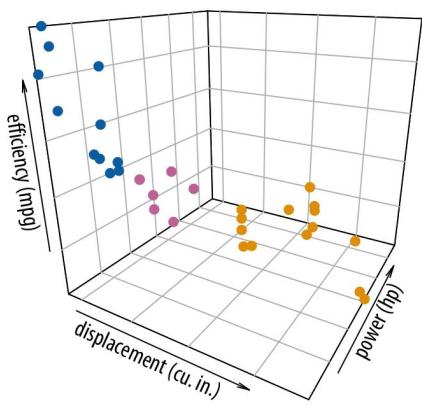
# Comments on Common Types of Figures

# 3D plots (avoid)

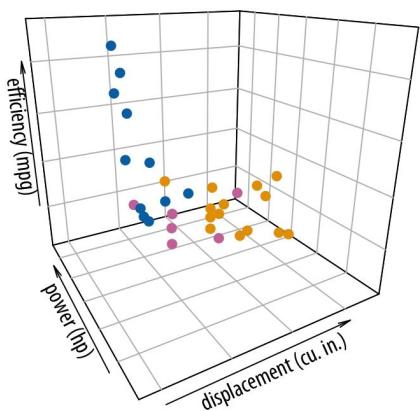
- Unless you are mapping geography or something like proteins in biology which we probably aren't



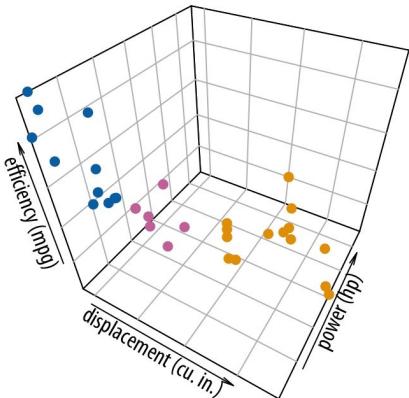
a



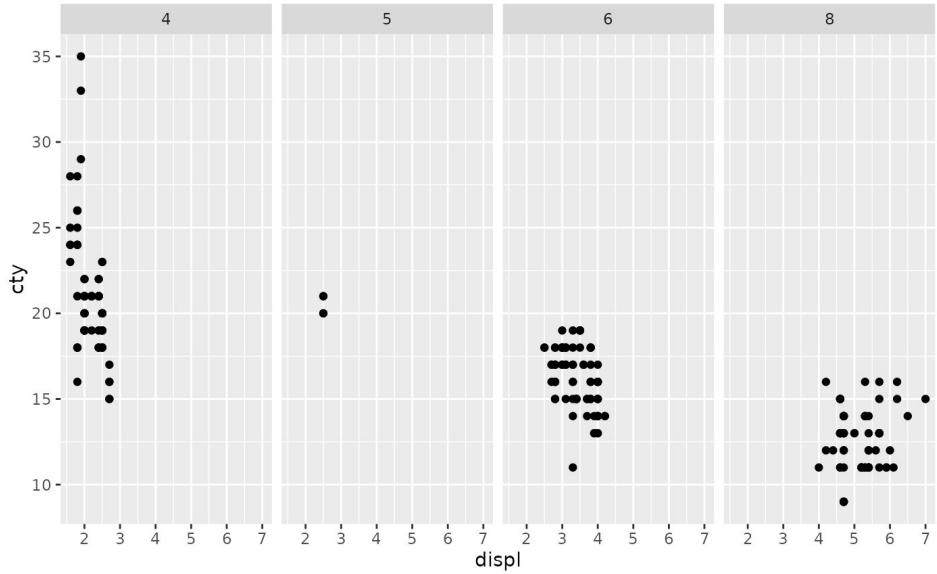
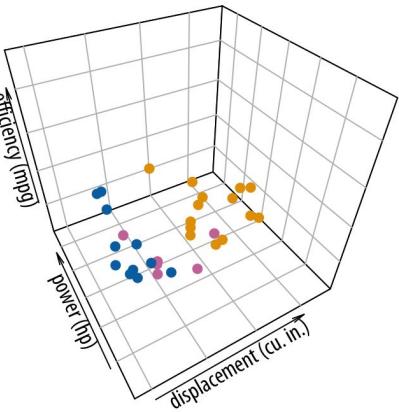
b



c



d

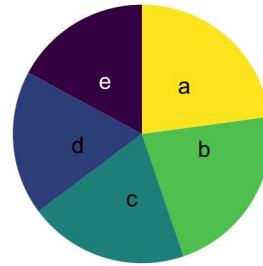
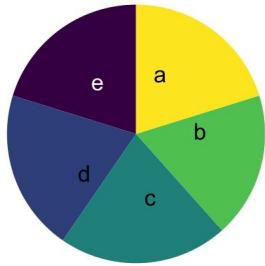
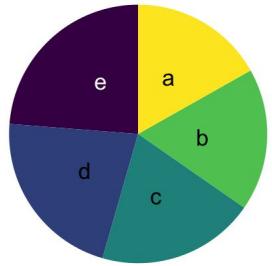


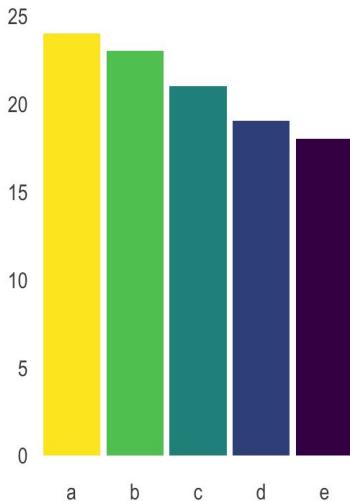
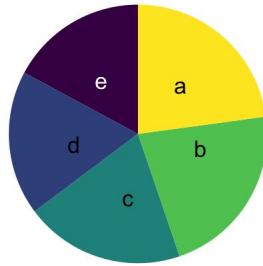
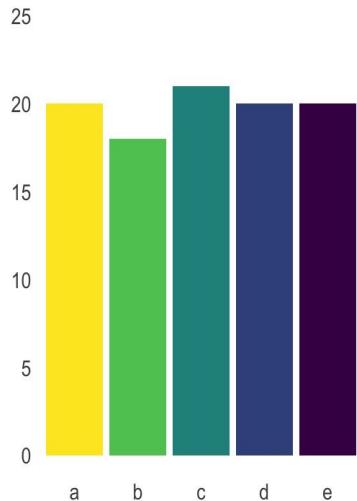
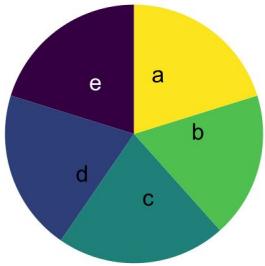
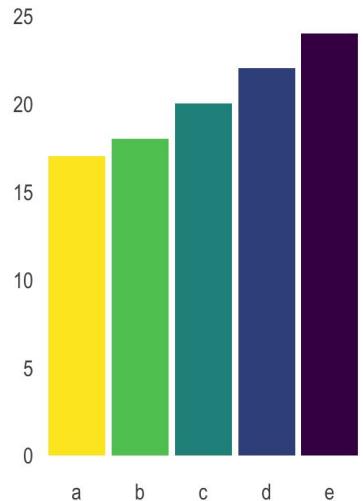
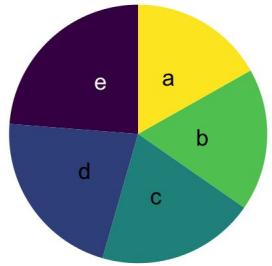
# Pie Charts (pretty much always avoid)

- *A table is nearly always better than a dumb pie chart; the only worse design than a pie chart is several of them, for then the viewer is asked to compare quantities located in spatial disarray both within and between charts ... Given their low density and failure to order numbers along a visual dimension, pie charts should never be used.* - Edward Tufte
- *...completely useless...* - Jacques Bertin
- *There is no data that can be displayed in a pie chart that cannot be displayed better in some other type of chart.* - John Tukey
- *Use these plots with caution - polar coordinates has major perceptual problems. The main point of these examples is to demonstrate how these common plots can be described in the grammar. Use with EXTREME caution.* [ggplot2 documentation for coord\\_polar\(\)](#)

# Pie Charts (pretty much always avoid)

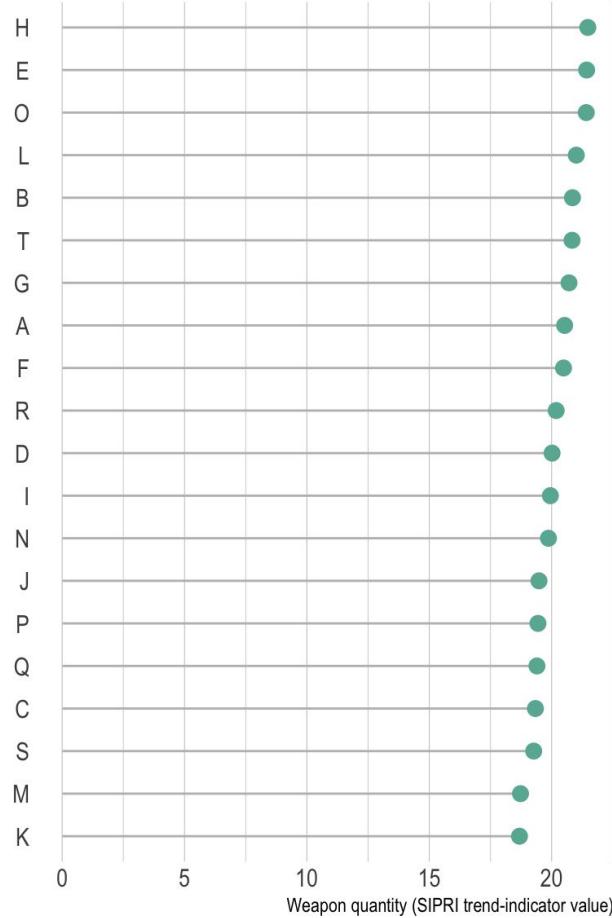
- Cleveland and McGill (1984) found that chart types based on length (such as dot charts and **bar charts**) were **read much more accurately than** chart types based on angle, area or volume (such as **pie charts** and three dimensional charts).

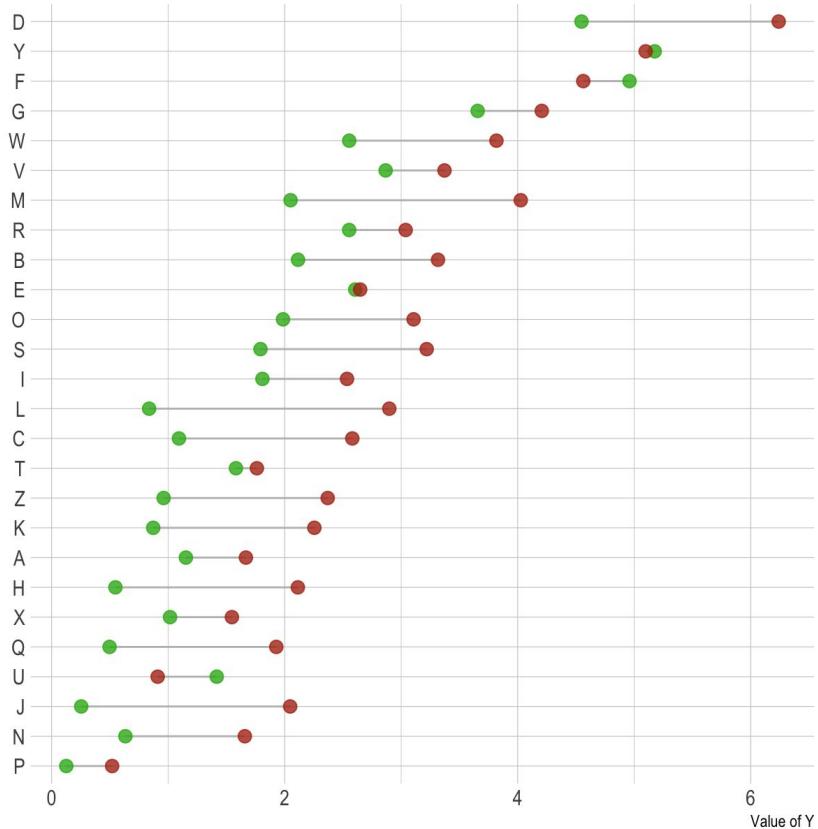




# Bar Charts (sometimes avoid)

- Good for showing percentages or counts (ie alternative to pie chart)
- Stacked bar charts aren't great
  - Furthermore, they [Cleveland and McGill (1984)] found that people perform substantially worse on stacked bar charts than on aligned bar charts, and that comparisons between adjacent bars are more accurate than between widely separated bars.
    - I still use them 😊
- Not good for showing uncertainty





Friends don't let friends make bar plots!

KICKSTARTER

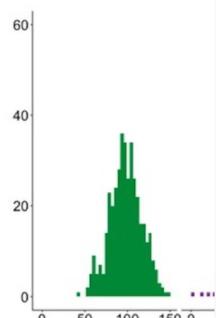
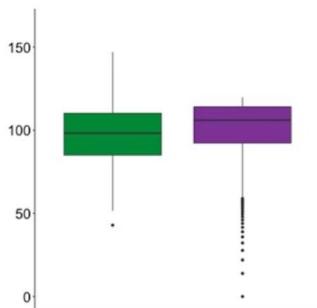
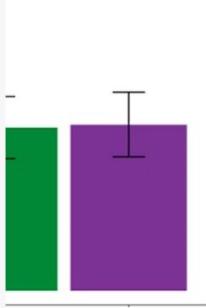
#barbarplots

## Friends don't let friends make bar plots.

These look the same!

Wait a minute...

Ooh!



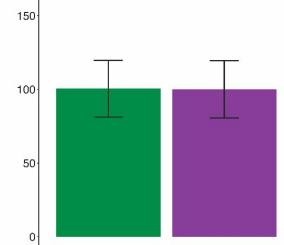
Bar plots fail to accurately represent the spread of data. Help us get the word out to journal editors to #barbarplots with t-shirts!

Created by  
Bar Bar Plots

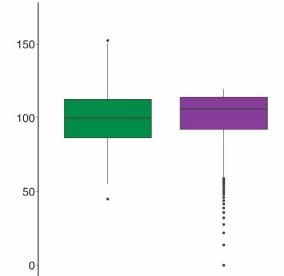
173 backers pledged €3,479 to help bring this project to life.

Last updated July 29, 2016

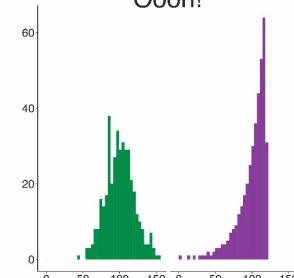
These look the same!



Wait a minute...



Ooh!

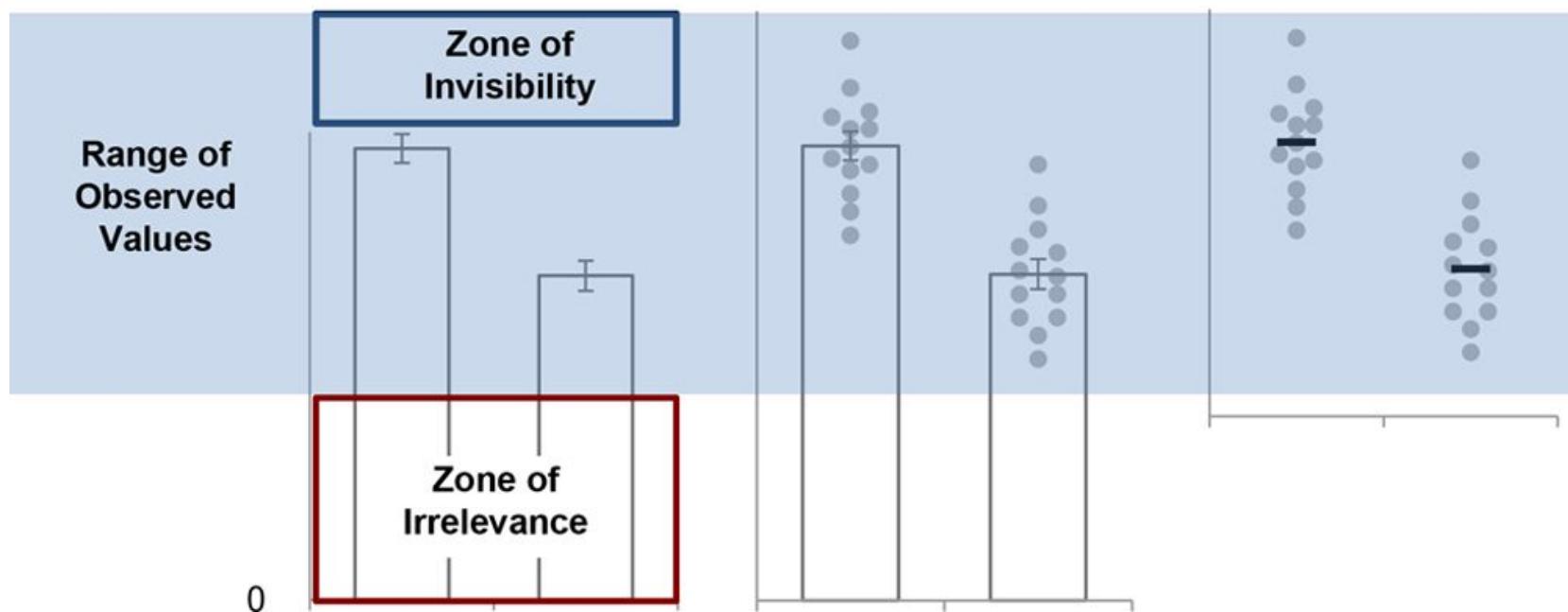


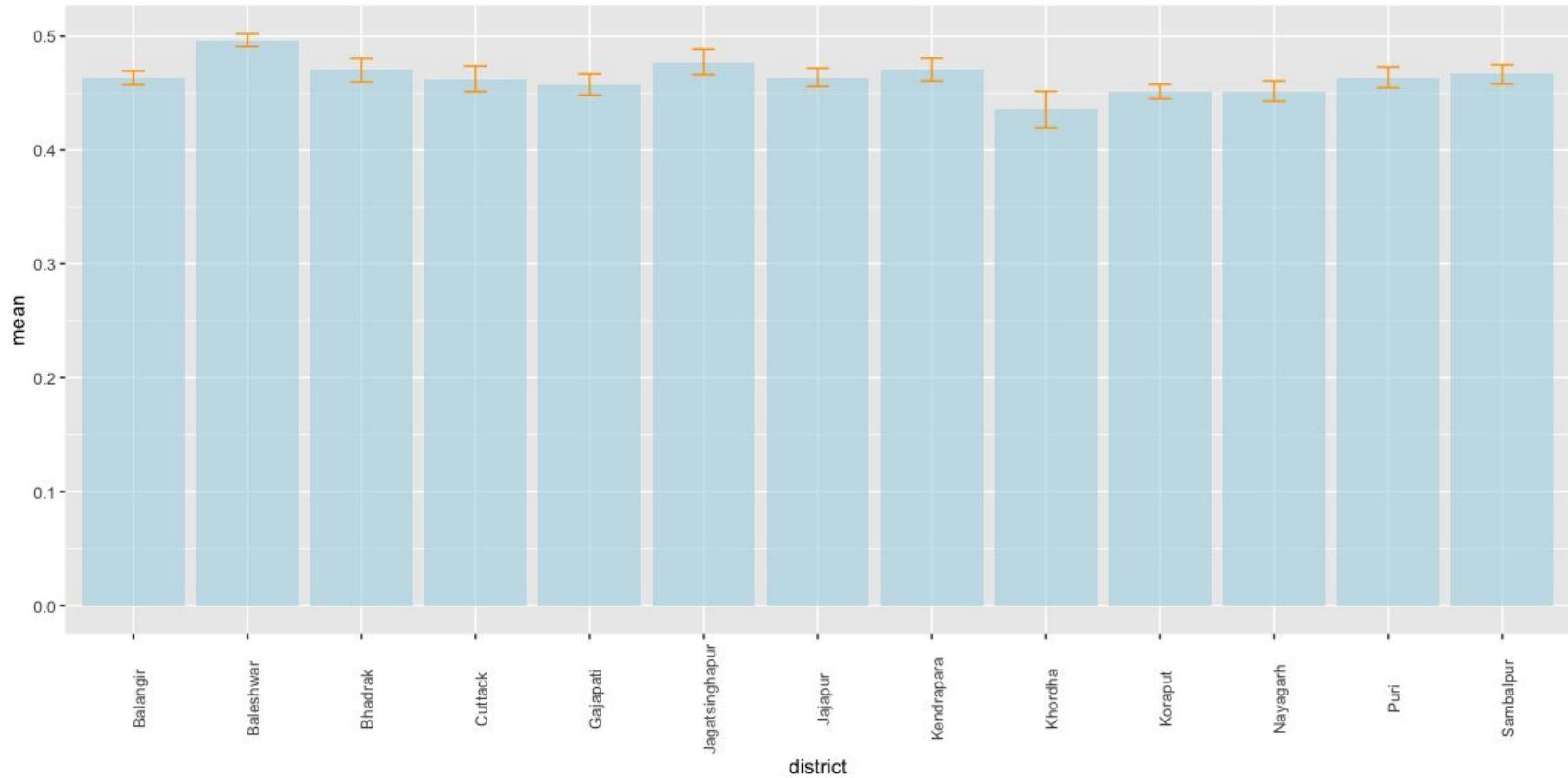
#barbarplots

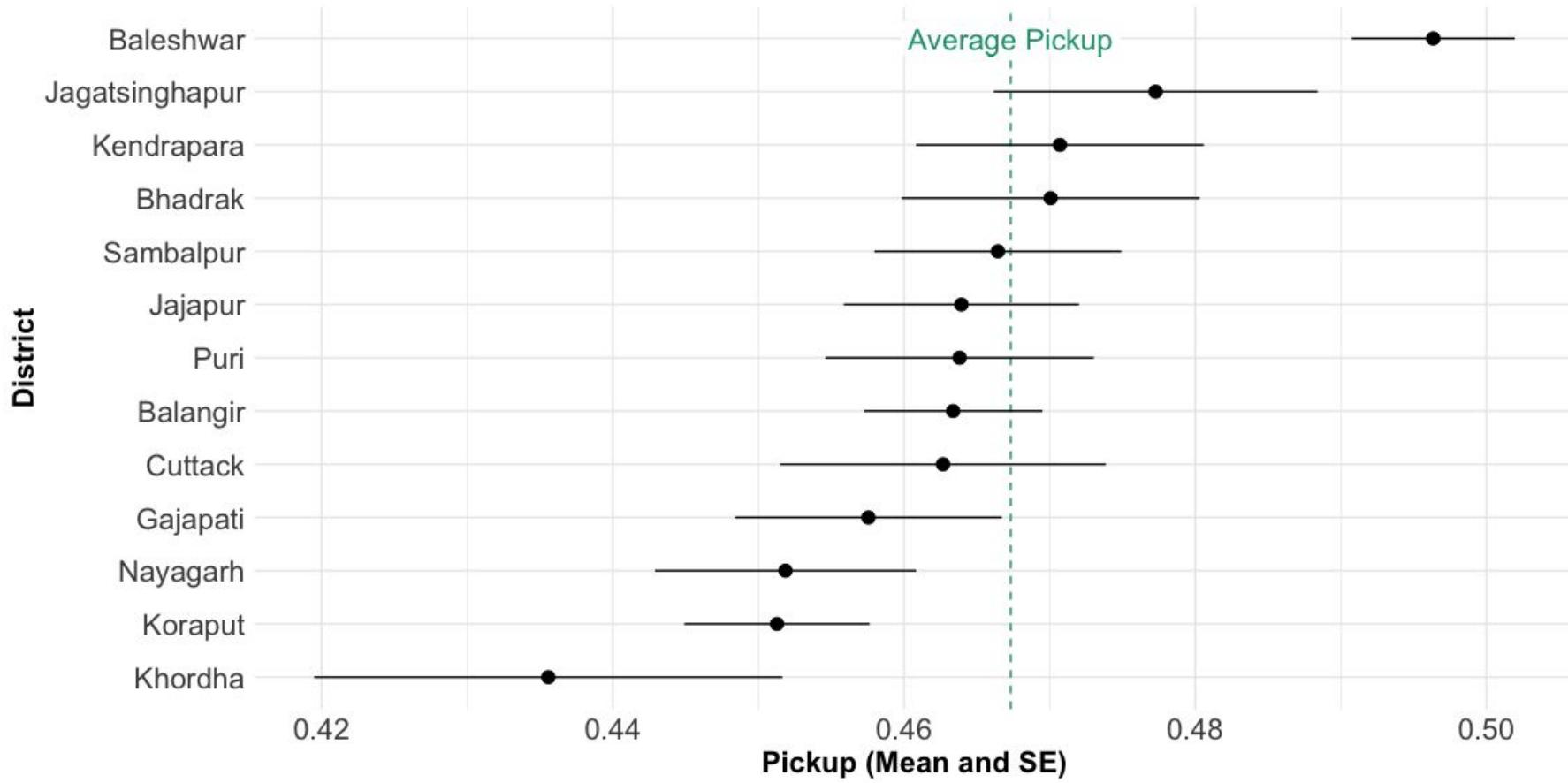
**A** Bar graph (mean  $\pm$  SE)

**B** Bar graph with dot plot

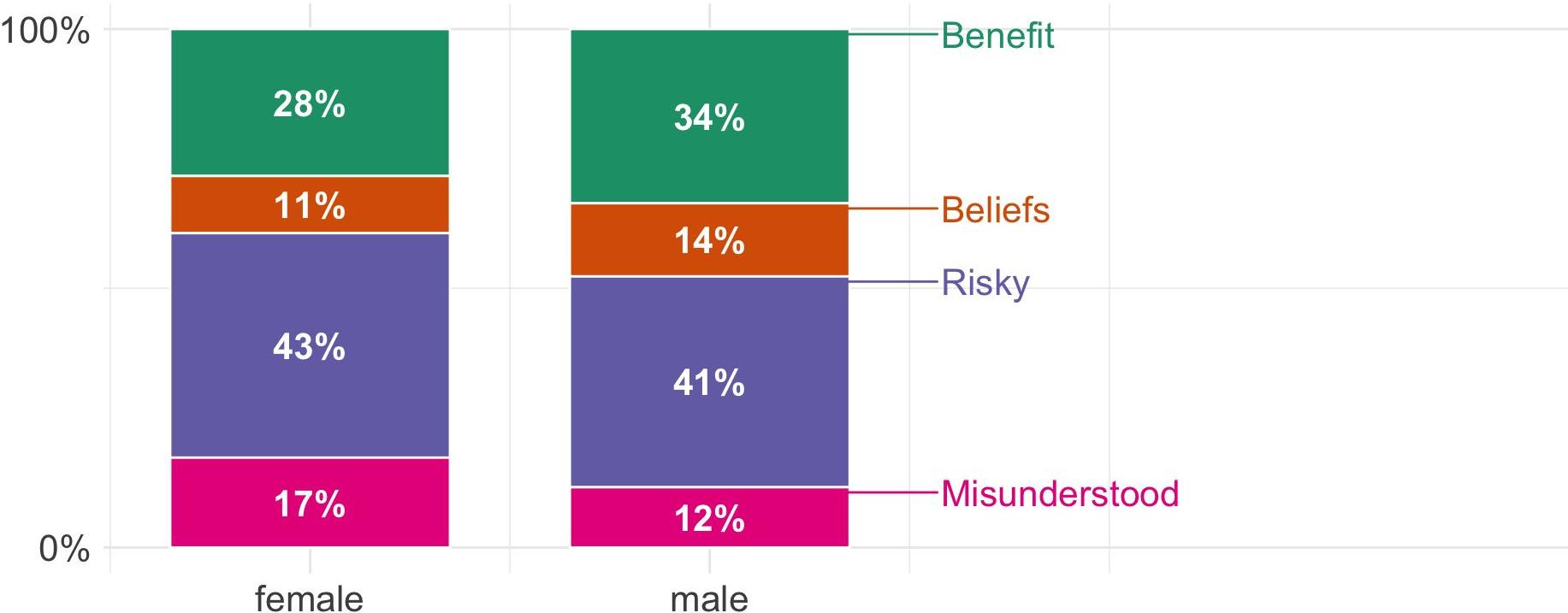
**C** Dot plot







# Motive Impediment by Gender



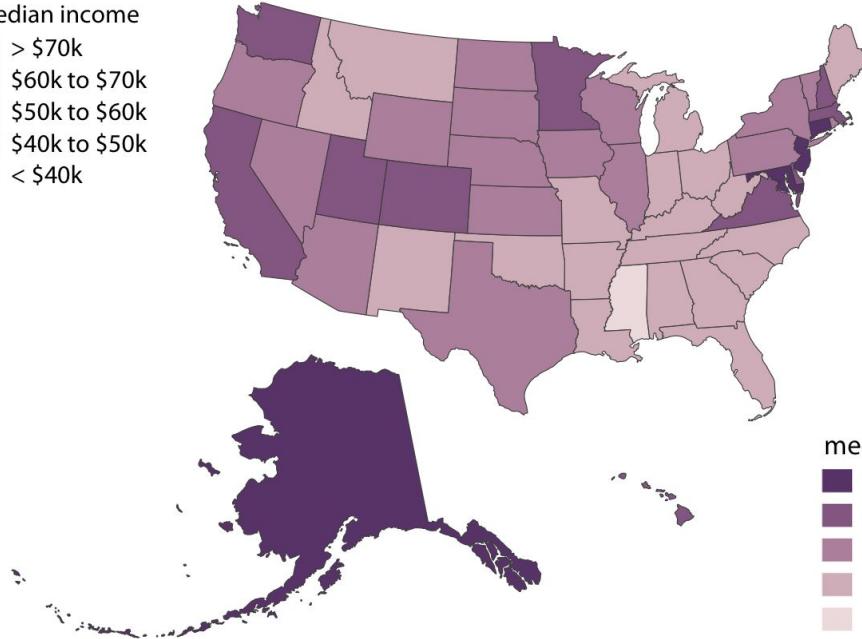
Number of Observations: 368

# Maps

- If you want to represent state-level data that does not depend on the geographic size of the state, use [statebins](#)
- There are "binned" maps for other countries and geographies with [geofacet](#)

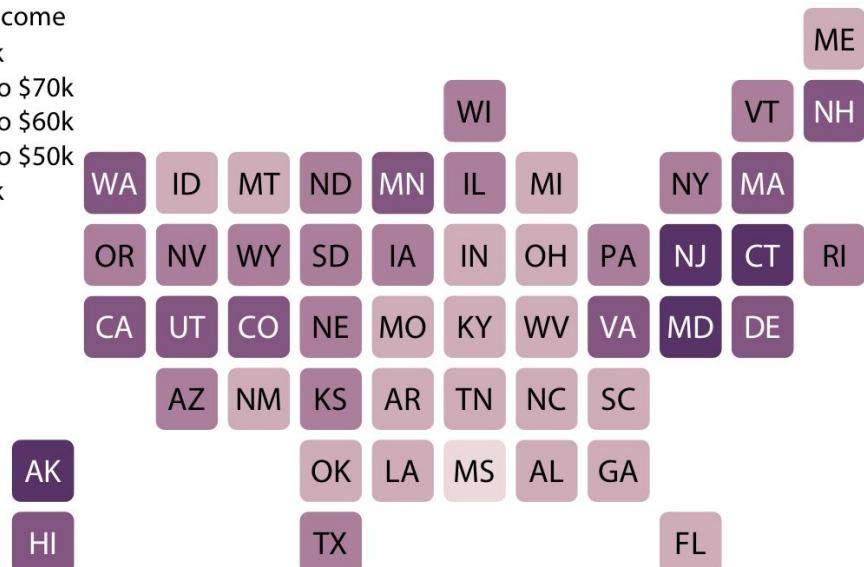
median income

- > \$70k
- \$60k to \$70k
- \$50k to \$60k
- \$40k to \$50k
- < \$40k

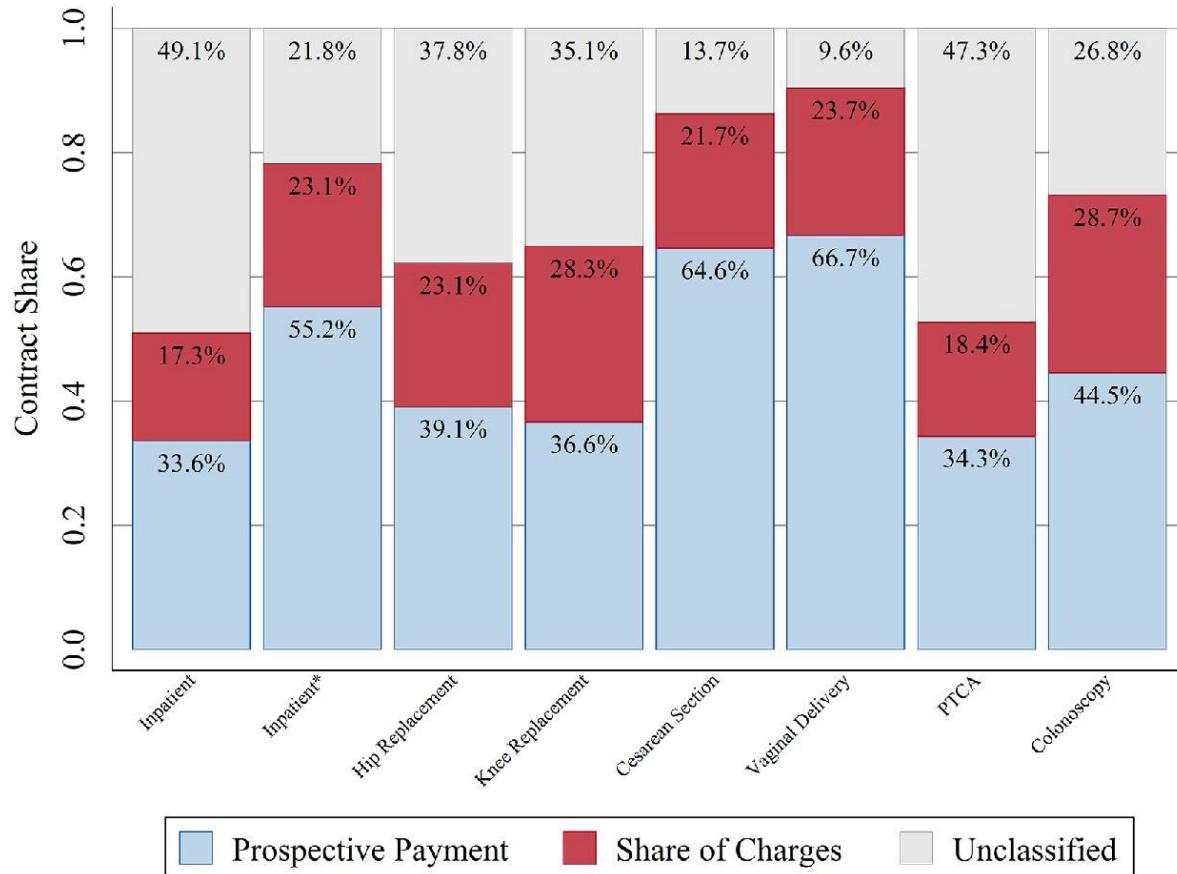


median income

- > \$70k
- \$60k to \$70k
- \$50k to \$60k
- \$40k to \$50k
- < \$40k



# Practice



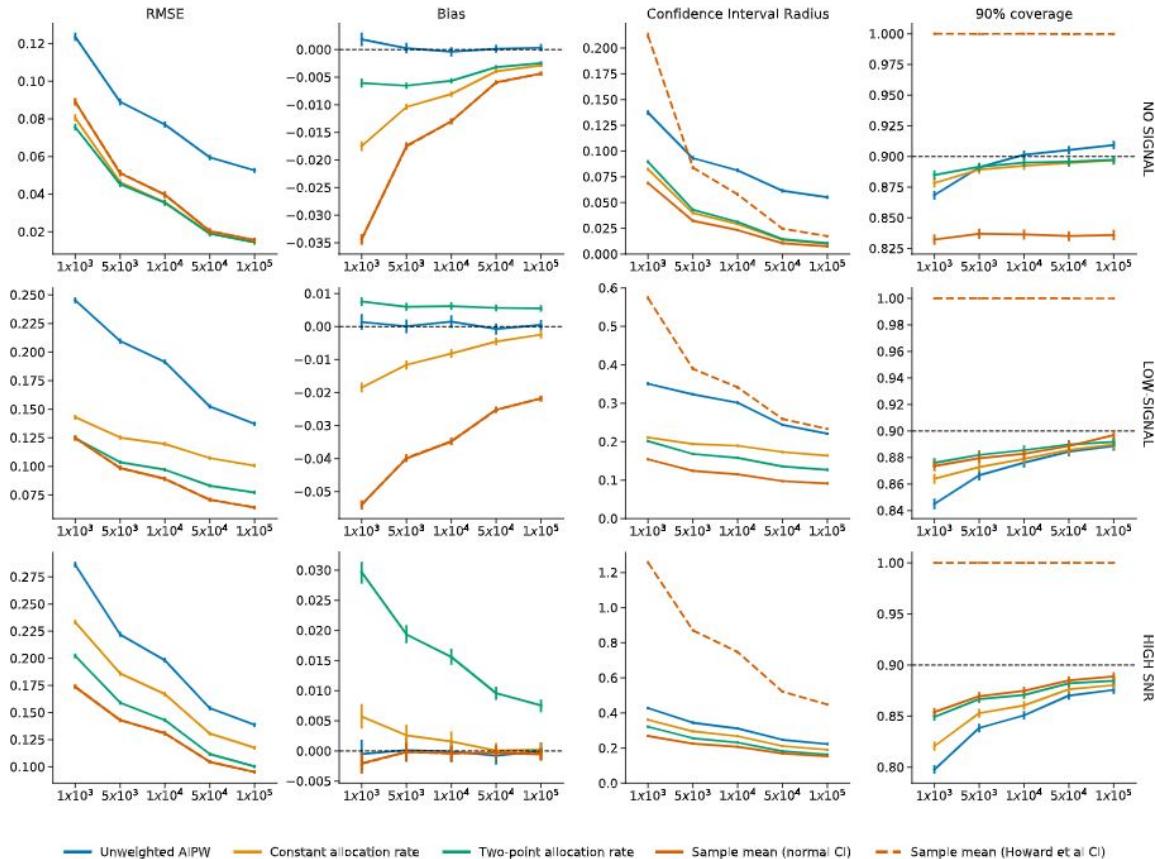
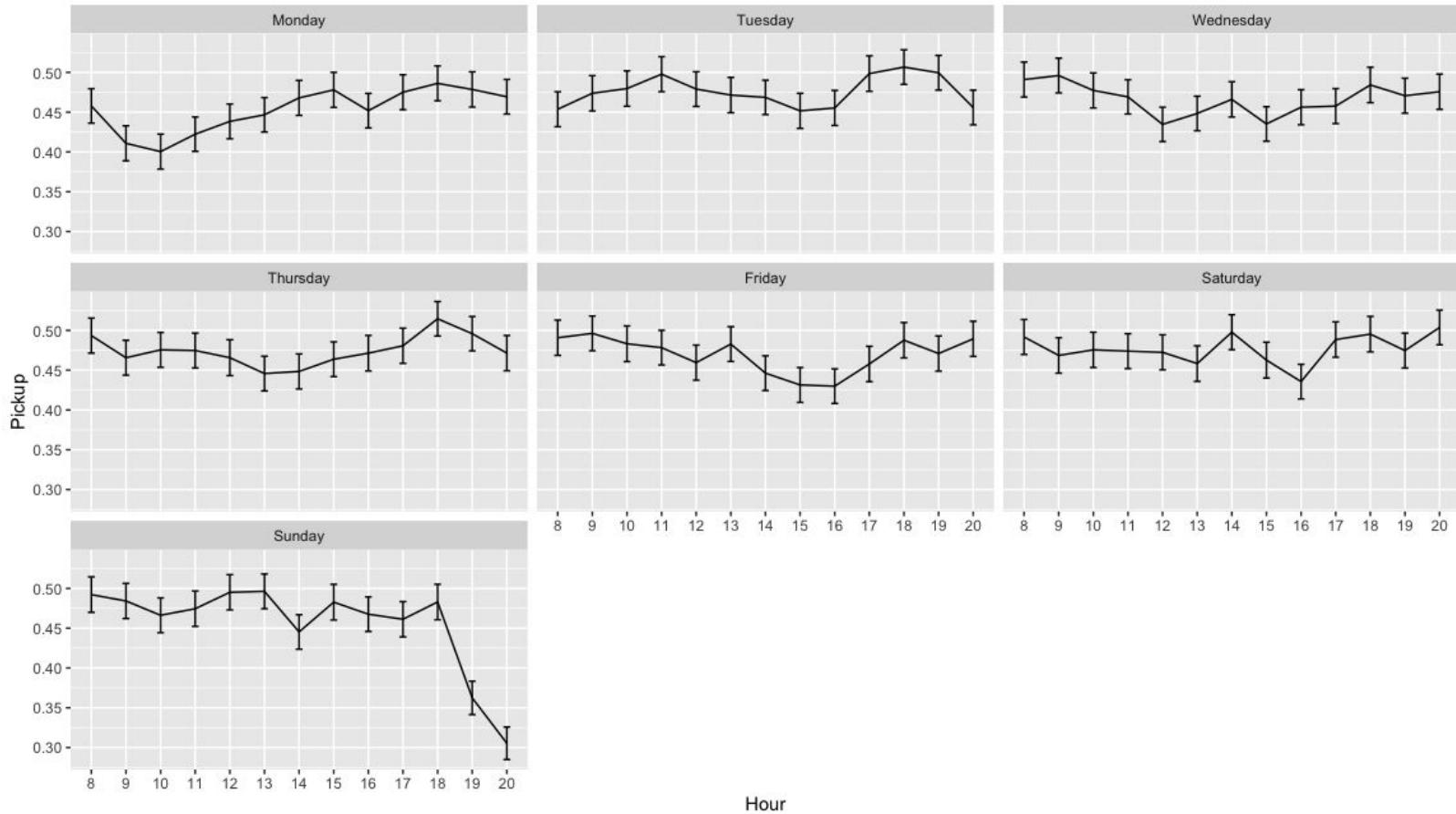
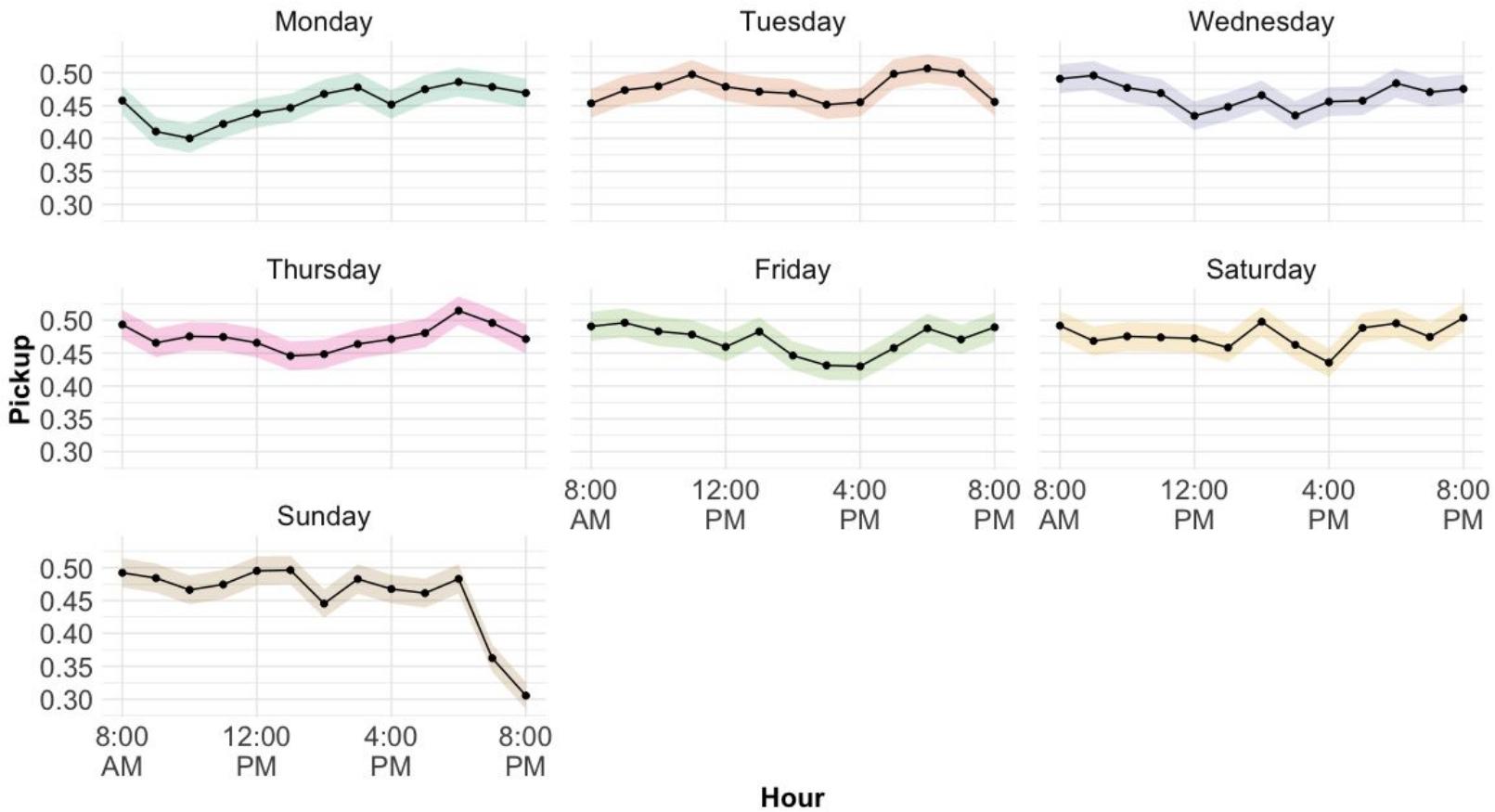
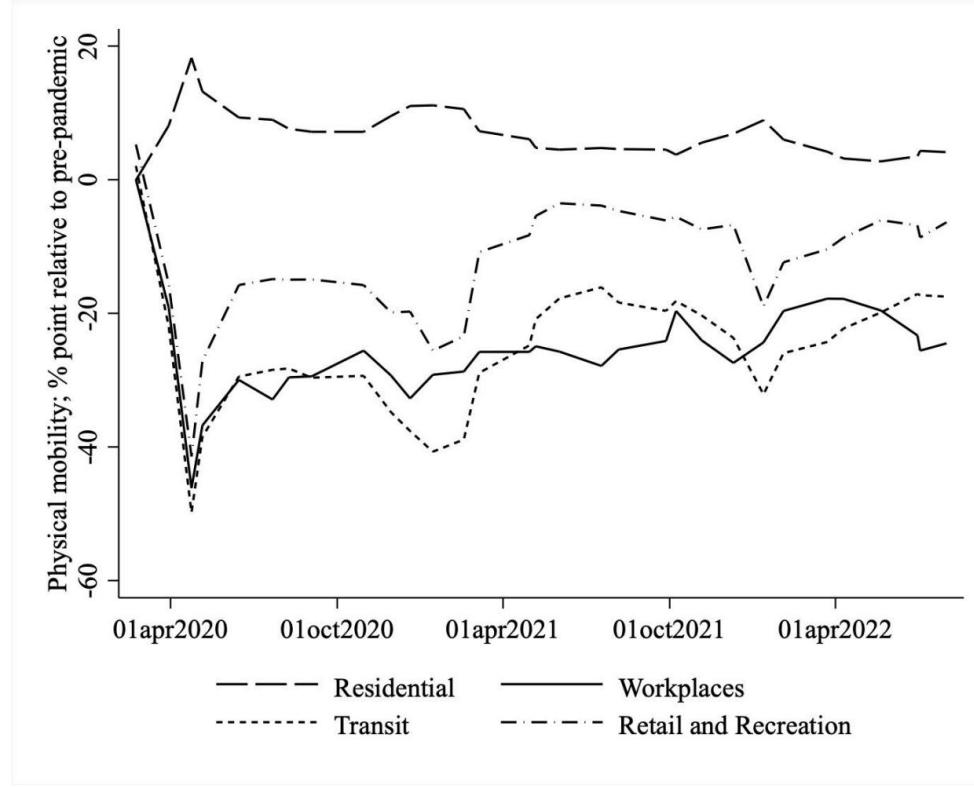


Figure 7: Estimates of the “bad” arm value  $Q(1)$  across for different experiment lengths.



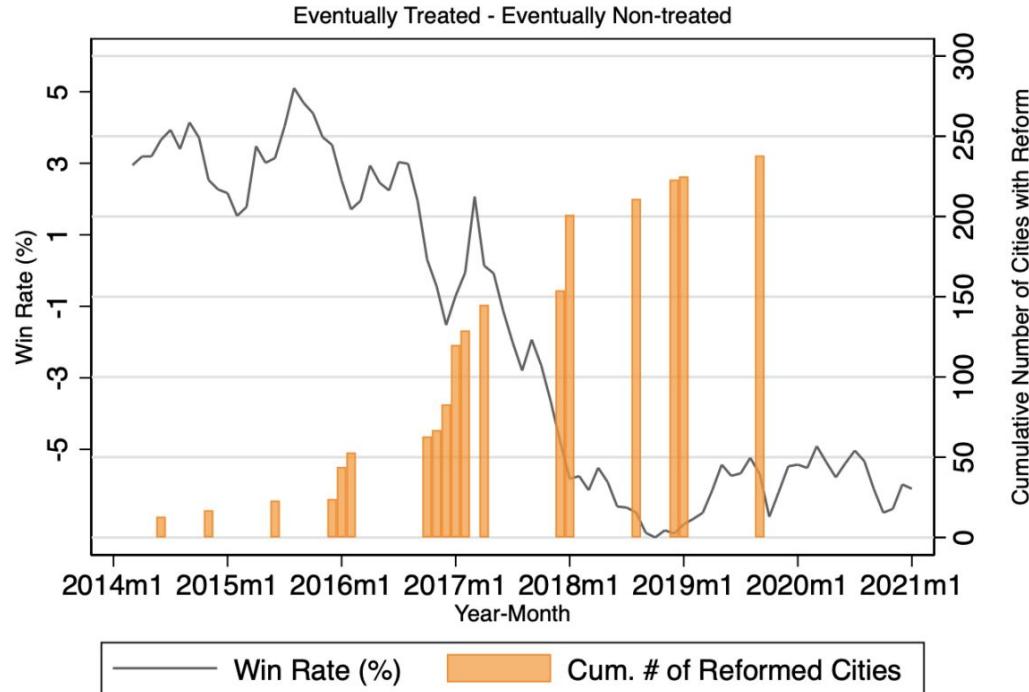


**Figure 1: Changes in physical mobility in the U.S. over the course of the Covid-19 pandemic**



Source: Google's COVID-19 [Mobility Reports](#) up to August 8, 2022 (accessed August 16).

## Judicial Independence Reform and the Win Rate of Local Defendants



(b) Reform Expansion & Local Defendants' Win Rate