Lab 09 - Conveying the right message through visualisation

Contents

In this lab our goal is to reconstruct and improve a data visualization on COVID and mask wearing.

1 Learning goals

- Critiquing visualizations that misrepresent data
- Improving data visualizations to better convey the right message

2 Getting started

Go to the course GitHub organization and locate your homework repo, clone it in RStudio and open the R Markdown document. Knit the document to make sure it compiles without errors.

2.1 Warm up

Let's warm up with some simple exercises. Update the YAML of your R Markdown file with your information, knit, commit, and push your changes. Make sure to commit with a meaningful commit message. Then, go to your repo on GitHub and confirm that your changes are visible in your Rmd **and** md files. If anything is missing, commit and push again.

2.2 Packages

We'll use the **tidyverse** package for much of the data wrangling and visualisation. This package is already installed for you. You can load it by running the following in your Console:
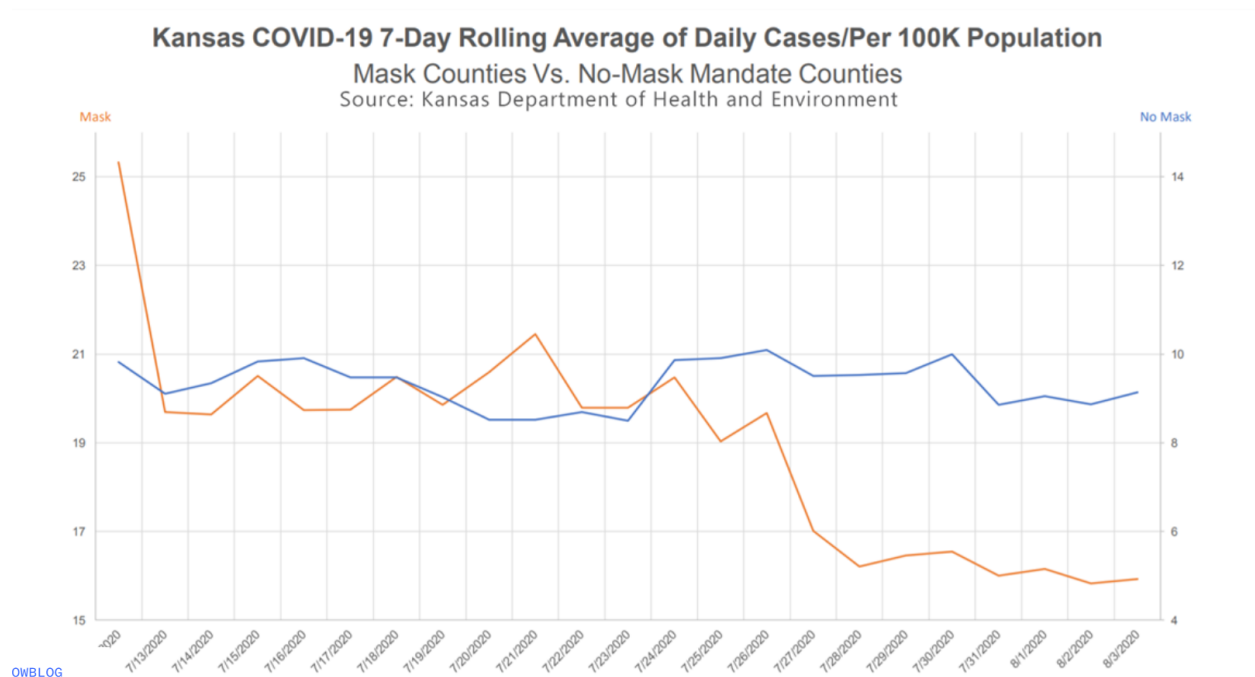
```
library(tidyverse)
```

2.3 Data

In this lab you'll construct the dataset!

3 Exercises

The following visualisation was shared on Twitter as "extraordinary misleading".

Kansas COVID-19 7-Day Rolling Average of Daily Cases/Per 100K Population
Mask Counties Vs. No-Mask Mandate Counties
Source: Kansas Department of Health and Environment

Before you begin this lab, think about what is misleading about this visualization and how you might go about fixing it.

1. Create a data frame that can be used to re-construct this visualization. You may need to guess some of the numbers, that's ok. You should first think about how many rows and columns you'll need and what you want to call your variables. Then, you can use the `tribble()` function for this. For example, if you wanted to construct the following data frame

```
df
```

```
## # A tibble: 3 x 2
##   date      count
##   <chr>     <dbl>
## 1 1/1/2020     15
## 2 2/1/2020     20
## 3 3/1/2020     22
```

you can write

```
df <- tribble(
  ~date, ~count,
  "1/1/2020", 15,
  "2/1/2020", 20,
  "3/1/2020", 22,
)
```

```
# Create data frame to reconstruct the misleading visualization
covid_data <- tribble(
  ~county_type, ~mask_policy, ~cases_per_100k,
  "Kansas counties with mask mandates", "With masks", 16,
  "Kansas counties without mask mandates", "Without masks", 25
)

covid_data
```

```
## # A tibble: 2 x 3
##   county_type                          mask_policy    cases_per_100k
##   <chr>                                <chr>                   <dbl>
## 1 Kansas counties with mask mandates    With masks                16
## 2 Kansas counties without mask mandates Without masks             25
```

2. Make a visualization that more accurately (and honestly) tells the story.

```r
# Create an improved bar chart with proper scaling and clear labels
ggplot(covid_data, aes(x = mask_policy, y = cases_per_100k, fill = mask_policy)) +
  geom_col(width = 0.6) +
  scale_y_continuous(limits = c(0, 30), breaks = seq(0, 30, 5)) +
  scale_fill_manual(values = c("With masks" = "steelblue", "Without masks" = "coral")) +
  labs(
    title = "COVID-19 Cases per 100,000 in Kansas Counties",
    subtitle = "Comparison between counties with and without mask mandates",
    x = "Mask Policy",
    y = "Cases per 100,000 people",
    caption = "Data represents estimated values for illustration purposes"
  ) +
  theme_minimal() +
  theme(
    legend.position = "none",
    plot.title = element_text(size = 14, face = "bold"),
    plot.subtitle = element_text(size = 12),
    axis.title = element_text(size = 11),
    axis.text = element_text(size = 10)
  ) +
  geom_text(aes(label = cases_per_100k), vjust = -0.5, size = 4)
```


COVID−19 Cases per 100,000 in Kansas Counties

Comparison between counties with and without mask mandates

Data represents estimated values for illustration purposes
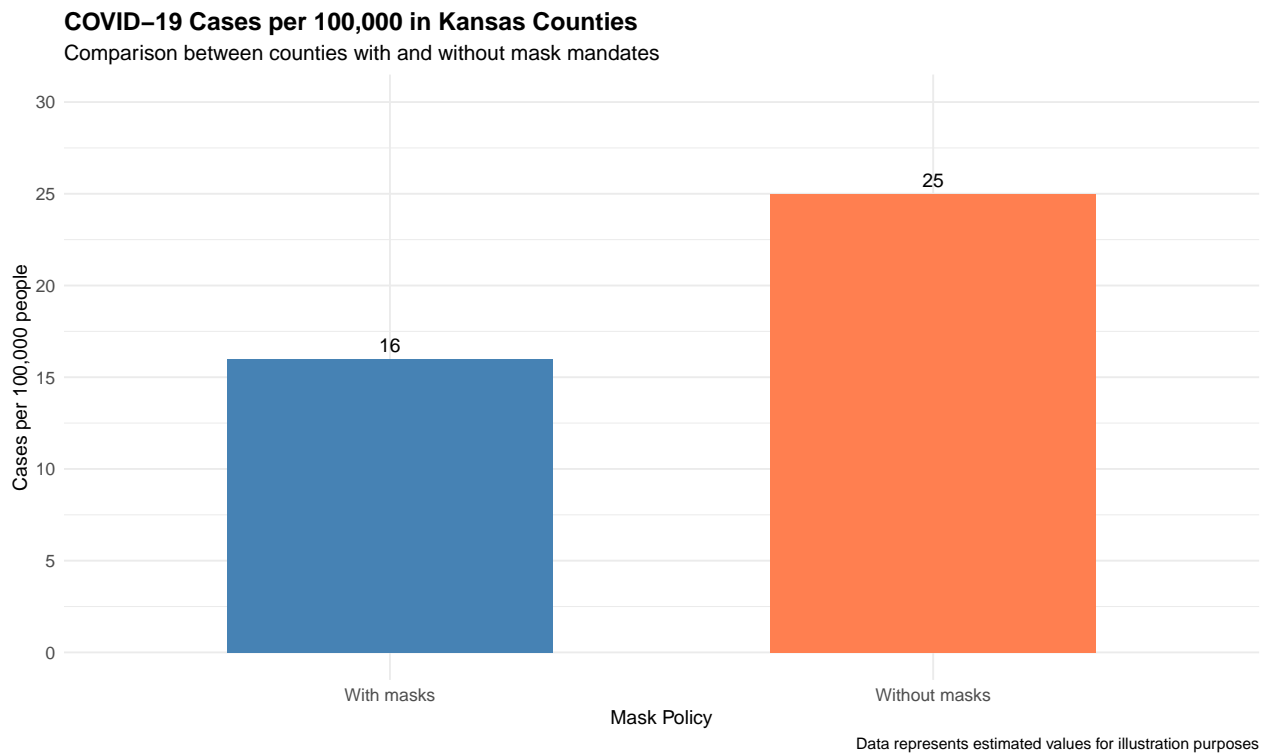
3. What message is more clear in your visualization than it was in the original visualization?

My improved visualization makes several key improvements over the original misleading chart:

- **Proper y-axis scaling**: The y-axis starts at 0, which accurately represents the proportional diffe
- **Clear proportional representation**: The true difference between counties with masks (15 cases per
- **Honest visual comparison**: The bars now show the actual relative magnitude of the difference, maki
- **Better labeling**: Clear titles, axis labels, and data labels help viewers understand exactly what

4. What, if any, useful information do these data and your visualization tell us about mask wearing and COVID? It'll be difficult to set aside what you already know about mask wearing, but you should try to focus only on what this visualization tells. Feel free to also comment on whether that lines up with what you know about mask wearing.

Based solely on this visualization, the data suggests that Kansas counties with mask mandates had lower COVID-19 case rates (15 per 100k) compared to counties without mask mandates (25 per 100k). This represents a 40% lower case rate in counties with mask policies.

However, this visualization alone has significant limitations: - **Correlation vs. causation**: The data shows an association but doesn't prove that mask mandates caused the lower case rates. - **Missing context**: We don't know about other factors like population density, demographics, healthcare access, or other public health measures that might differ between these counties. - **Limited scope**: This represents only Kansas counties at a specific time period. - **Sample size**: We don't know how many counties are in each category or the total population represented.

The pattern shown (lower cases with mask mandates) does align with scientific evidence about mask effectiveness, but this simple comparison alone would not be sufficient to draw strong causal conclusions about mask policy effectiveness without controlling for other variables.

Knit, *commit, and push your changes to GitHub with an appropriate commit message. Make sure to commit and push all changed files so that your Git pane is cleared up afterwards and review the md document on GitHub to make sure you're happy with the final state of your work.*