

TDU

R FOR DATA VISUALIZATIONS

Day 2

May 3, 2023



Public Health
Agency of Canada

Agence de la santé
publique du Canada

Canada

WELCOME BACK!

Yesterday we reviewed:

- Graphing using base R and ggplot
- Grammar of graphics
- Modifying your dataset to facilitate data visualization
- Custom data visualizations

OVERVIEW

Today we will cover:

- Best practices in data visualization
- Application of graphic design practices in R
- Data visualization and potential for harm to small and vulnerable communities

A background image showing two women in a professional setting. One woman is pointing at a laptop screen while the other looks on. The image is overlaid with a dark blue semi-transparent filter. A bright blue curved shape is at the bottom right.

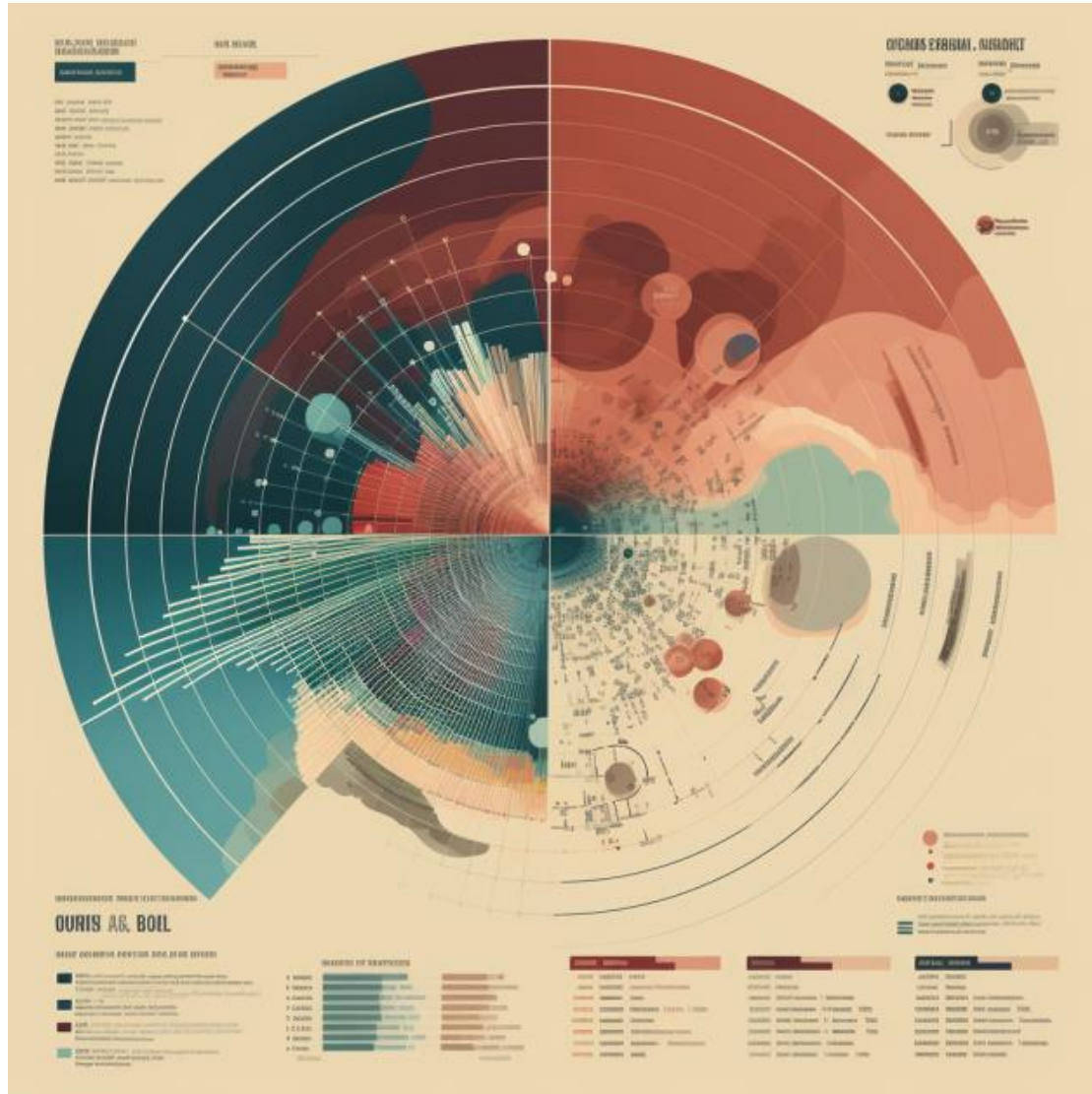
APPLYING BEST PRACTICES IN DATA VISUALIZATION

CLARITY OF MESSAGE

- What am I trying to say?
- To whom am I saying it?
- Why am I saying it?



Image Generated by MidJourney AI



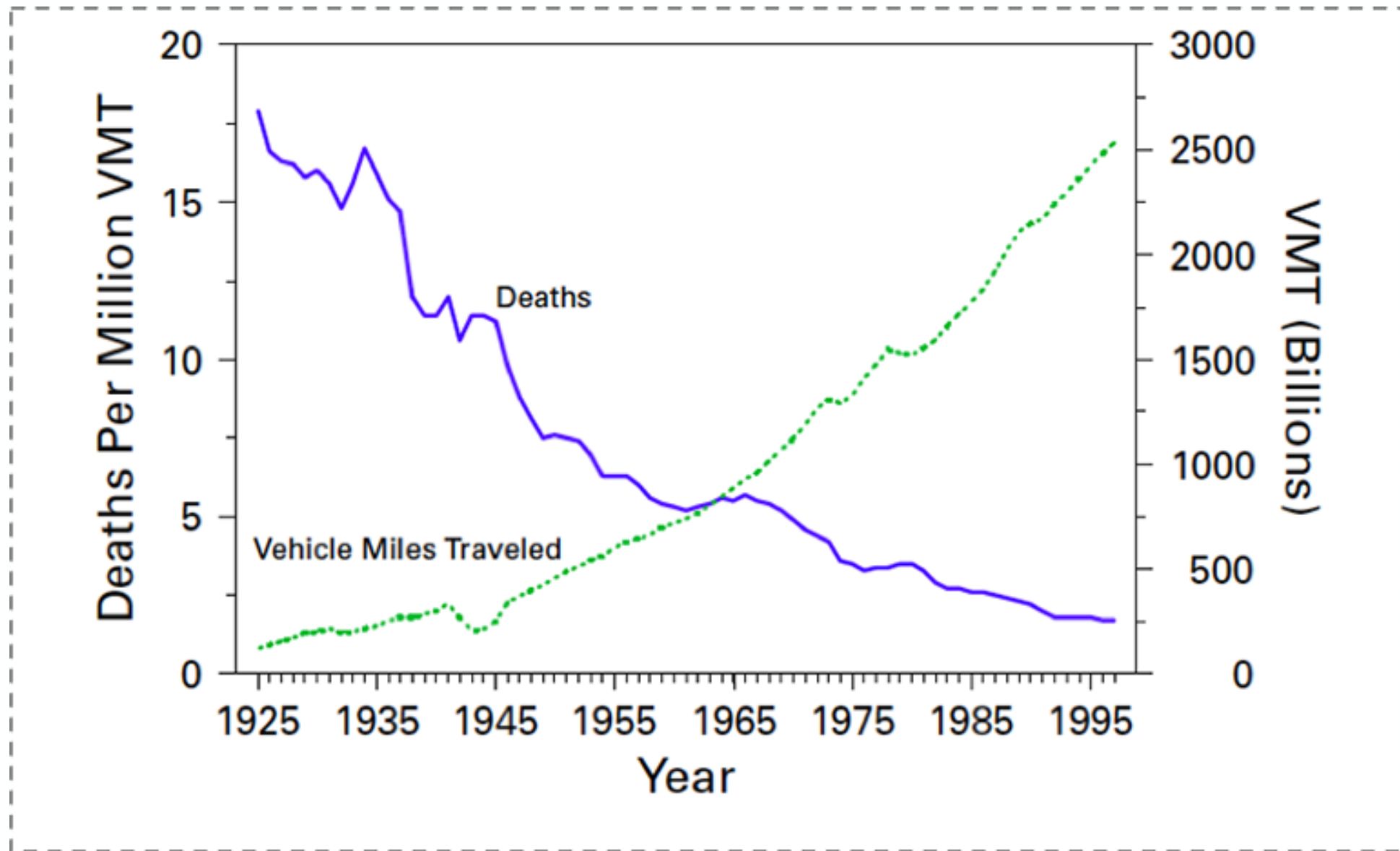
BEST PRACTICES

- To achieve clarity of message, consider:
 - Simplicity
 - Visual consistency and layout
 - Readability
 - Scale

SIMPLICITY

Avoid unnecessary or complex elements in your figures.

Limit visualizations to one concept each.



Motor-vehicle-related deaths per million vehicle miles traveled (VMT) and annual VMT, by year – United States, 1925– 8
1997. Motor vehicle safety: A 20th century public health achievement. MMWR 1999; 48(18): 369–374.

VISUAL CONSISTENCY

Colour, Contrast
and Consistency

Emphasis, balance
and hierarchy

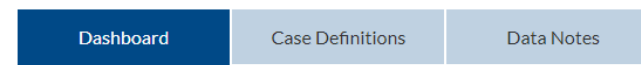
Layout and
negative space.

Vaccination Coverage Progress in BC by age and by Health Authority, 26 Mar. 2023



Chronic Disease Dashboard

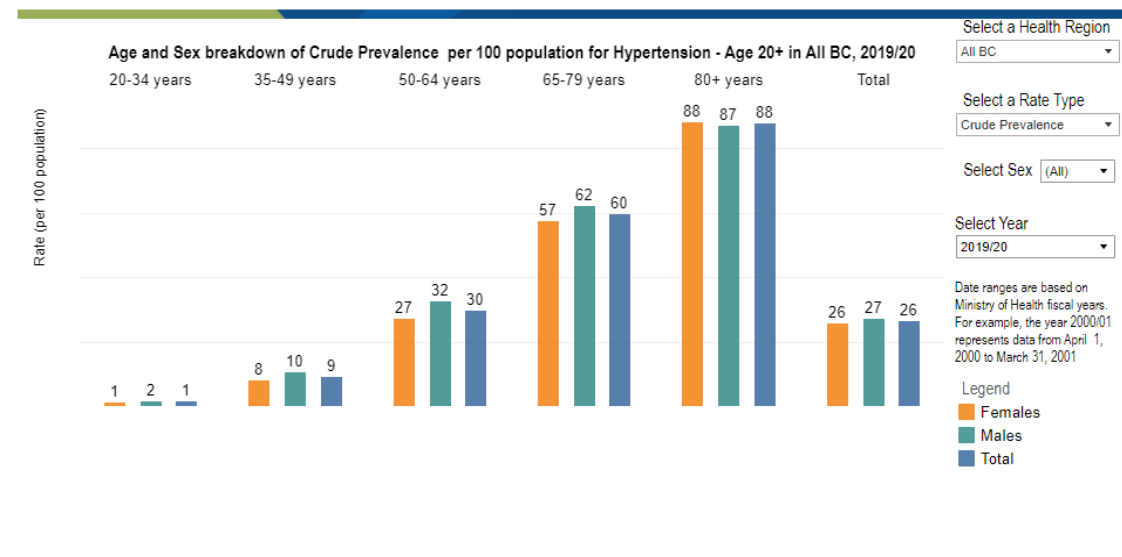
This interactive tool provides summary statistics on variety of non-communicable diseases and conditions in BC.



Select from the tools on the right to change diseases and display options. Click on the tabs to view different arrangements of the data. To learn how to save charts and data tables, [review our quick guide](#). Note that the tool may not display on older web browsers.



Visit the BC Observatory for Population & Public Health >





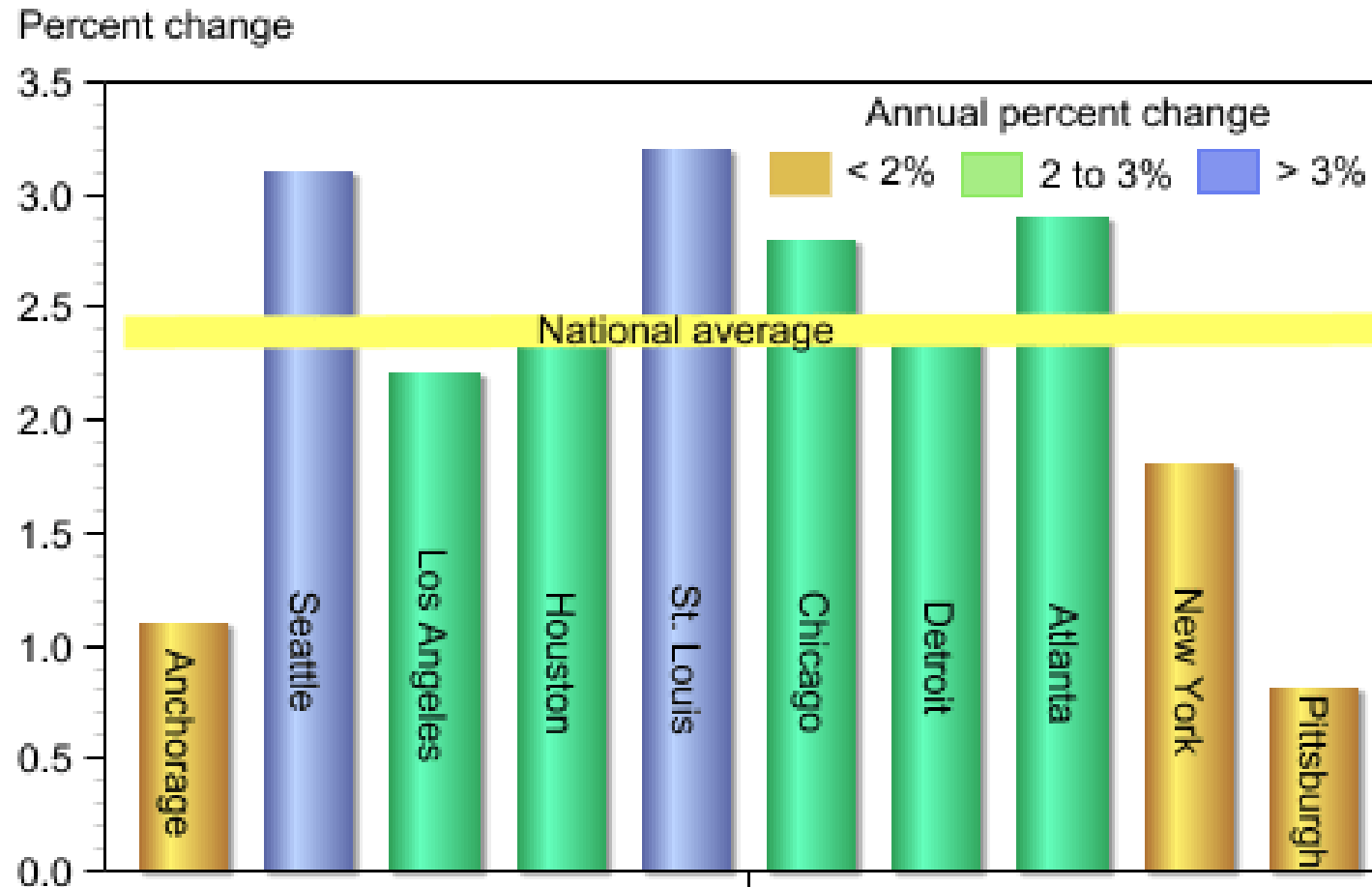
TUBIKSTUDIO.COM

**If everything yells for your
viewer's attention, nothing
is heard.**

Aarron Walter, "Design for Emotion"

Annual grocery store inflation by city, 2014

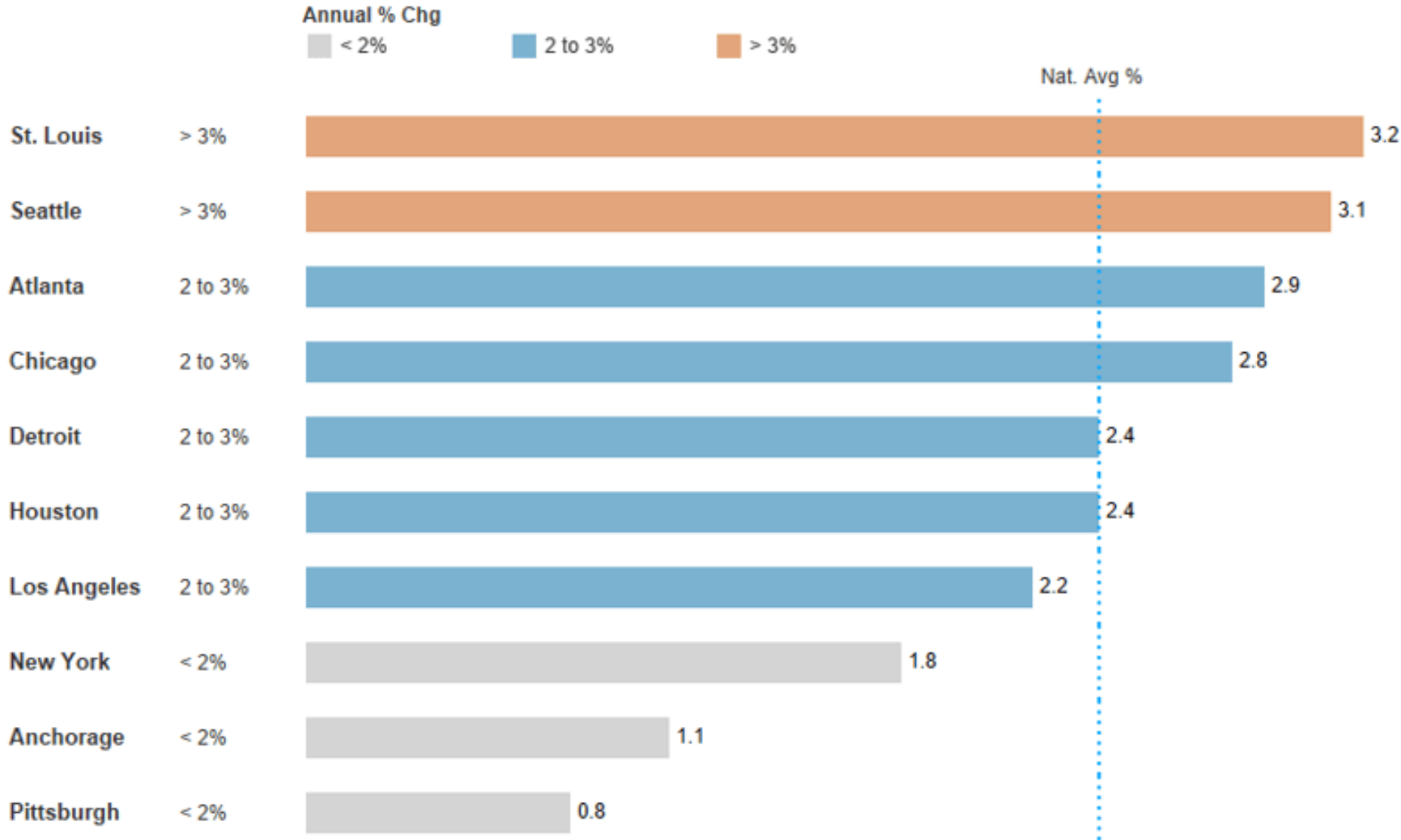
In 2014, food price inflation was higher in Seattle, St. Louis, Chicago, and Atlanta than in other metropolitan areas

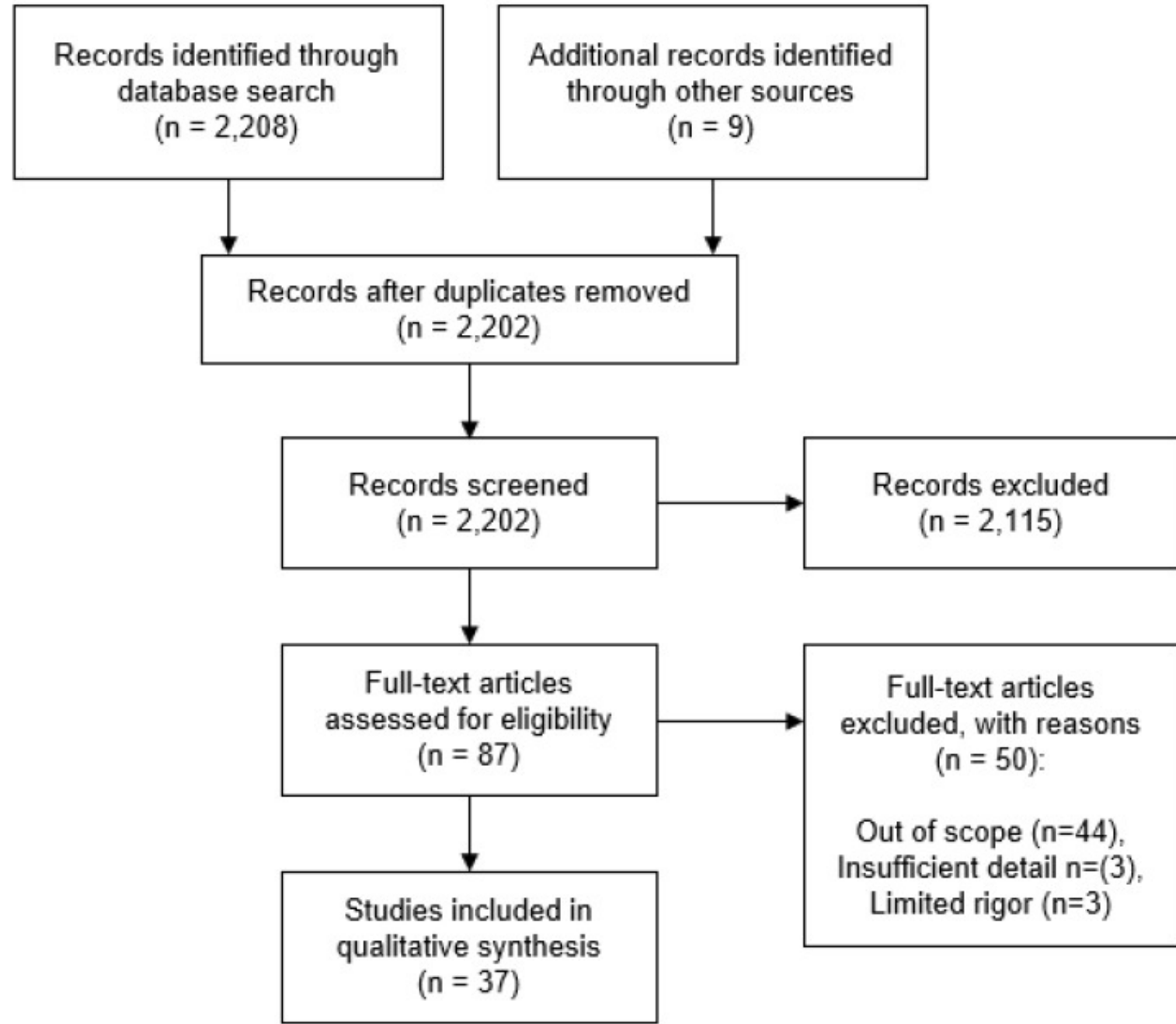


Source: Calculated by ERS, USDA, using Bureau of Labor Statistics (BLS) data.

2014 Annual grocery store inflation by city

Food price inflation was higher in Seattle, St. Louis, Chicago, and Atlanta than in other metropolitan areas





READABILITY OF TEXT

Simple, non-decorative

Consistent stroke width

Wider horizontal proportions

Distinct forms for each character

Limited bold weighting

USE THIS

Arial

Times New Roman

Verdana

Calibri

NOT THIS

Britannic

Comic Sans MS

Jokerman

Papyrus

SCALE

- Consider how the scale of your typography will impact your visualization's readability, hierarchy and intuitive perception of your graph.
- Also consider how the scale of your visual elements can contribute or detract from your key message.

Within the **Categories**
the **code** highlights the
diversity of opinions

632 total statements (Wave 6)

The inner circle represents the
category for each statement.
The corresponding outer circle
represents the codes within the
category



ACCESSIBILITY

Plain language

Colour contrast and
palette selection

Appropriate font

Visually consistent layout



APPLYING BEST PRACTICES FOR DATA VISUALIZATION IN R

The "How" Section

THEMES!

Themes: What are they?

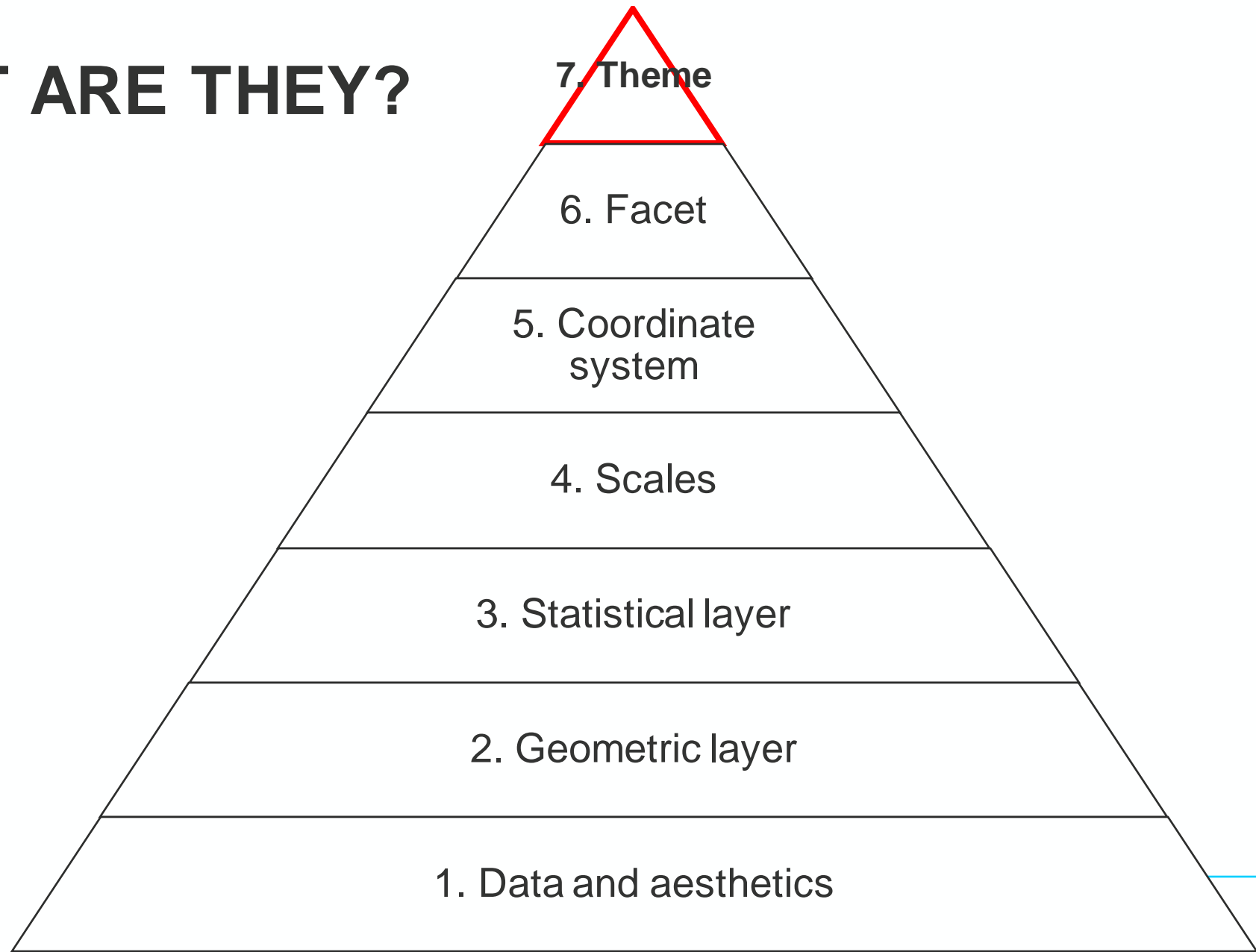
Function: `theme()`

Built-in themes and
downloading themes

Customizing themes

THEMES: WHAT ARE THEY?

- Themes are the collection of elements that allow you to adjust the non-data bits of your plot: e.g., background, font, etc.



THEME()

- In ggplot2, the primary means of modifying a plot's theme is by using the `theme()` function.
- It allows you to override the default theme elements, or even change, on the fly, theme elements on an existing theme.

```
ggplot(hospital_data, aes(x = date_onset)) +
```

← The data to be visualized

```
theme_minimal() +
```



```
theme(
```

```
  plot.title = element_text(face = "bold", size = 12),
```

```
  legend.background = element_rect(
```

```
    fill = "white",
```

```
    linewidth = 4,
```

```
    colour = "white"
```

```
),
```

```
  legend.position = "right",
```

```
  axis.ticks = element_line(colour = "grey70", linewidth = 0.2),
```

```
  panel.grid.major = element_line(colour = "grey70", linewidth = 0.2),
```

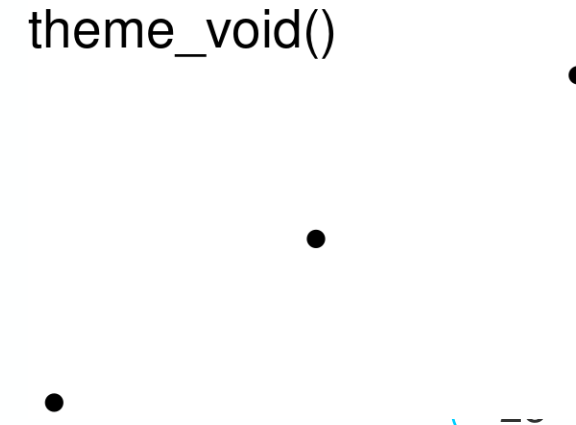
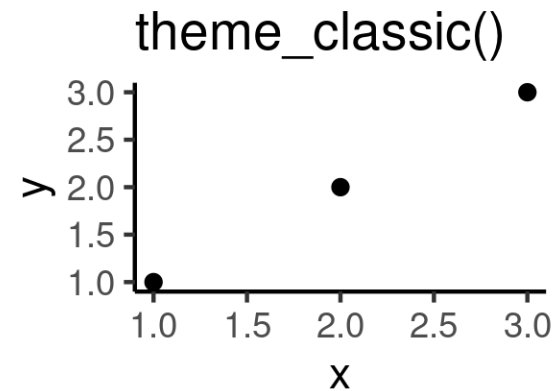
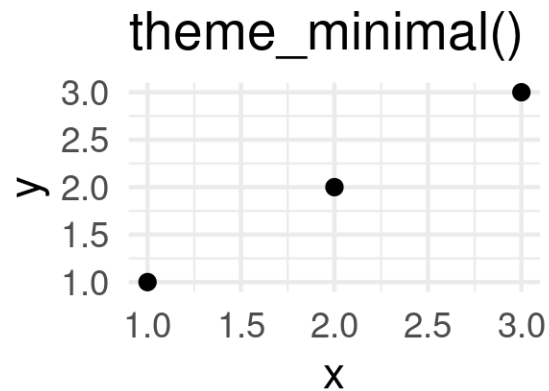
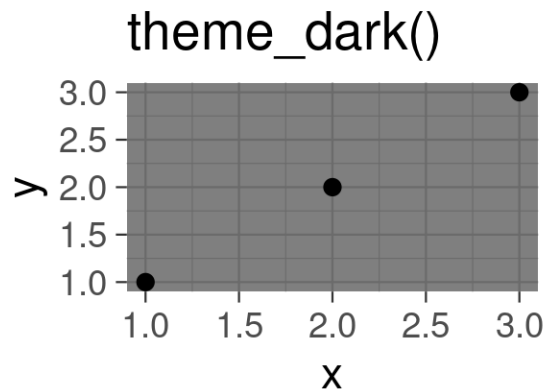
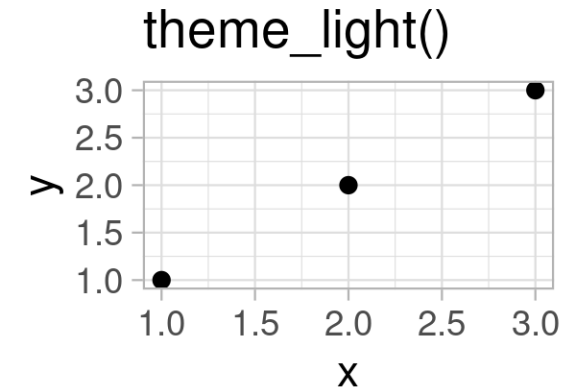
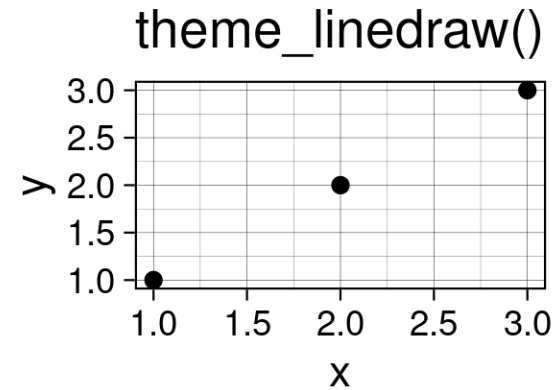
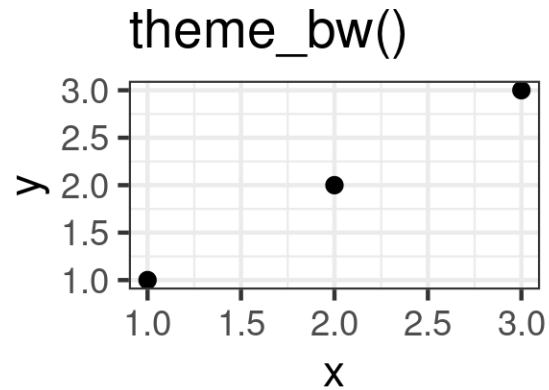
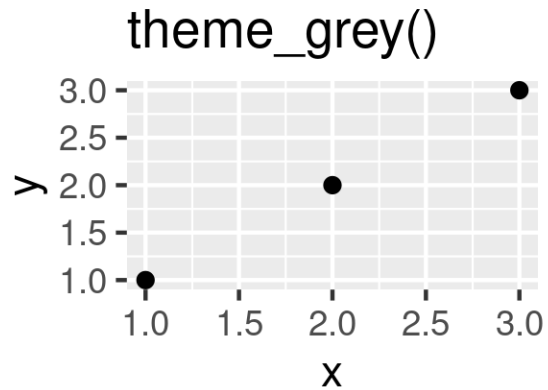
```
  panel.grid.minor = element_blank()
```

```
)
```

Application of an existing theme
to minimize background
elements

BUILT-IN THEMES

- ggplot2 contains a few utilitarian basic themes:



INVOKING BUILT-IN THEMES

- Calling a built-in theme is as simple as adding `theme_name()` to your plot object!

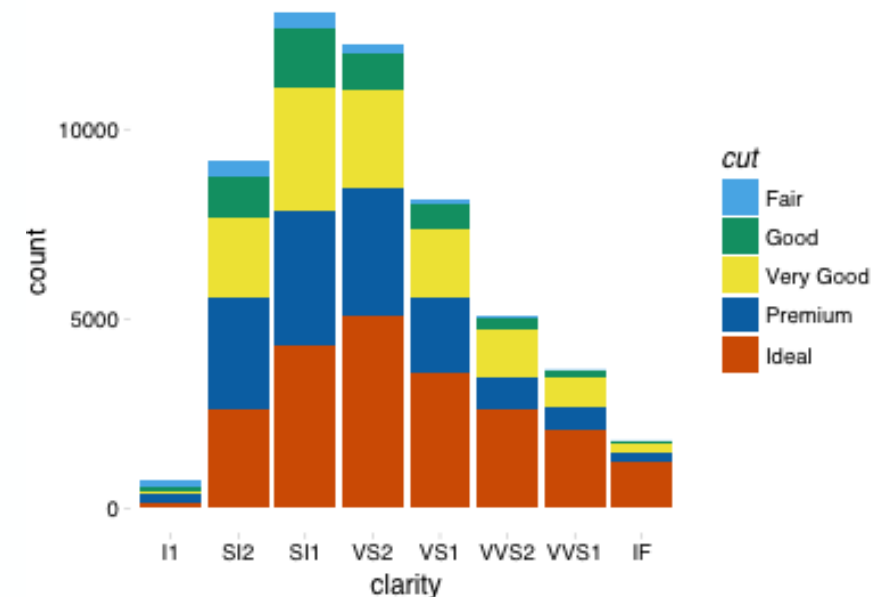
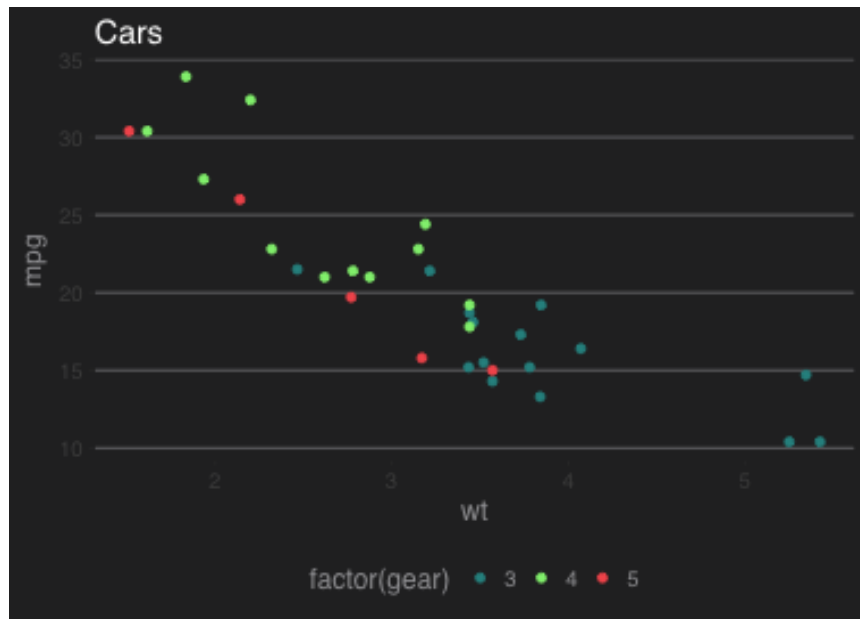
```
ggplot(hospital_data, aes(x = date_onset)) +
```

```
  theme_light() +
```

```
  theme( legend.position = "right",  
    ... )
```

INSTALLING ADDITIONAL THEMES

- You can also get more themes simply by installing a package that contains themes, or even copying and pasting the theming statements from other plots you've worked on.
- For example, the ggthemes package contains many different themes such as:



CUSTOMIZING THEMING

- We'll walk through a few examples now on some common plot “elements” that can be customized inside the `theme()` function and how to do so.
- Reminder: the ggplot documentation is your best friend, as is Google and Stack Overflow!
- Disclaimer: The following slides are shamelessly stolen from <https://ggplot2-book.org/polishing.html#modifying-theme-components>

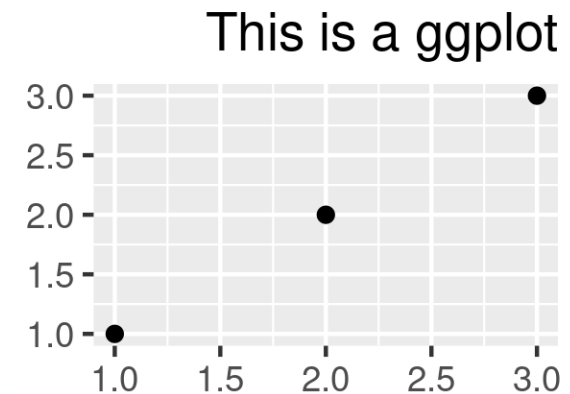
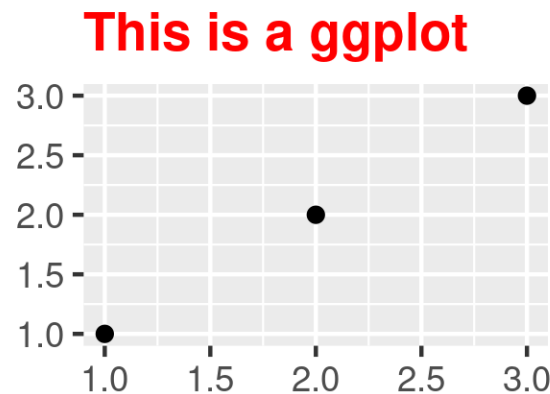
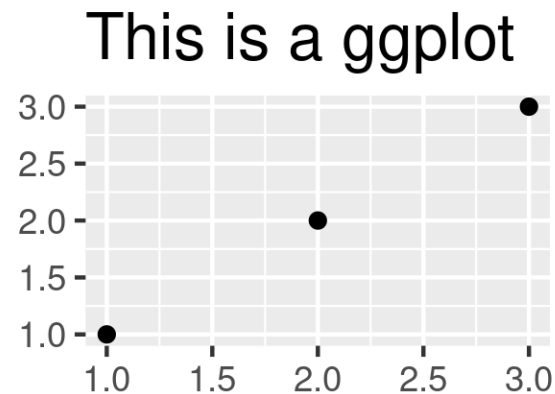
ELEMENT_TEXT

```
plot_t <- plot + labs(title = "This is a ggplot") + xlab(NULL) + ylab(NULL)
```

```
plot_t + theme(plot.title = element_text(size = 16))
```

```
plot_t + theme(plot.title = element_text(face = "bold", colour = "red"))
```

```
plot_t + theme(plot.title = element_text(hjust = 1))
```

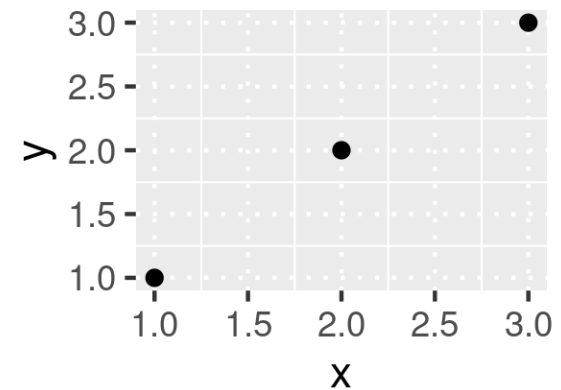
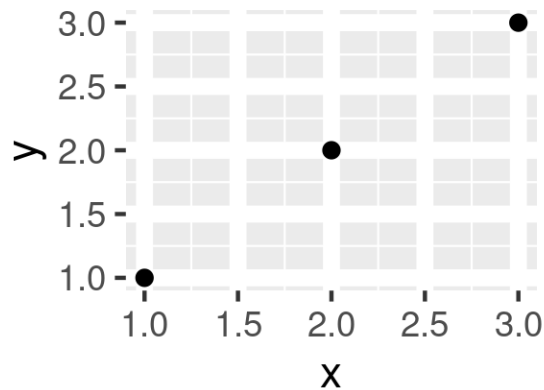
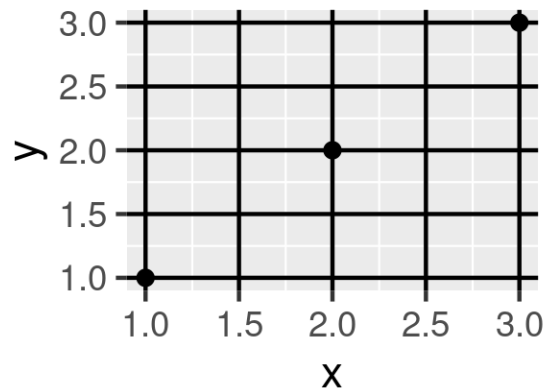


ELEMENT_LINE

`plot + theme(panel.grid.major = element_line(colour = "black"))`

`plot + theme(panel.grid.major = element_line(linewidth = 2))`

`plot + theme(panel.grid.major = element_line(linetype = "dotted"))`

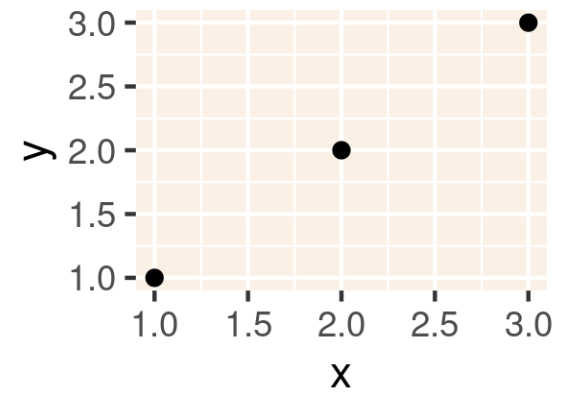
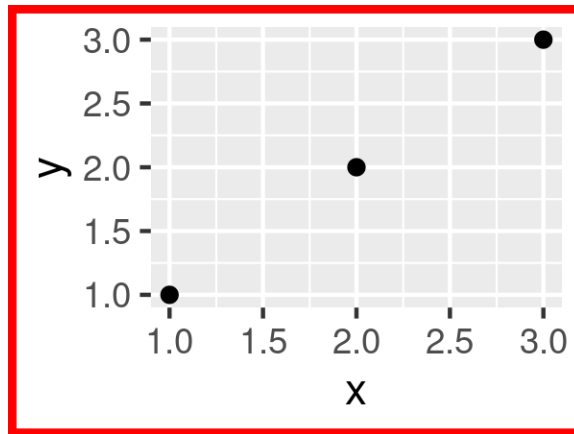
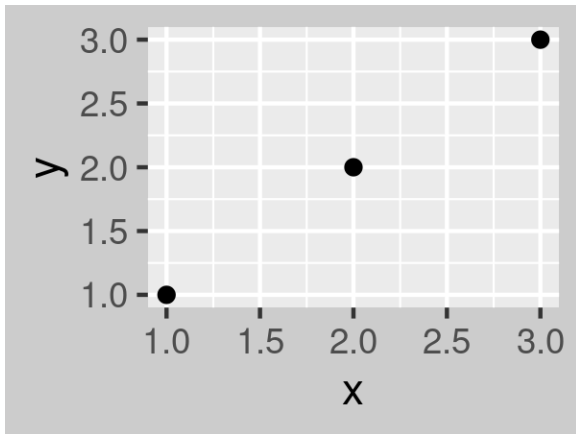


ELEMENT_RECT

```
plot + theme(plot.background = element_rect(fill = "grey80", colour = NA))
```

```
plot + theme(plot.background = element_rect(colour = "red", linewidth = 2))
```

```
plot + theme(panel.background = element_rect(fill = "linen"))
```

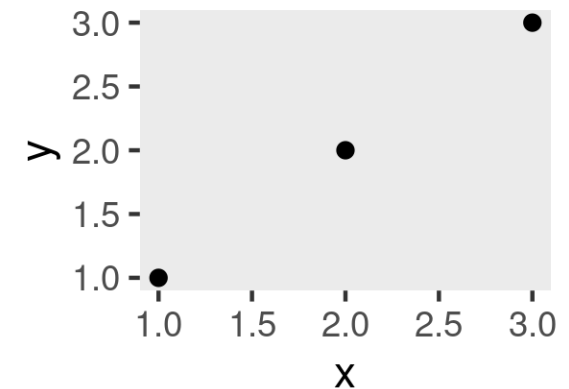
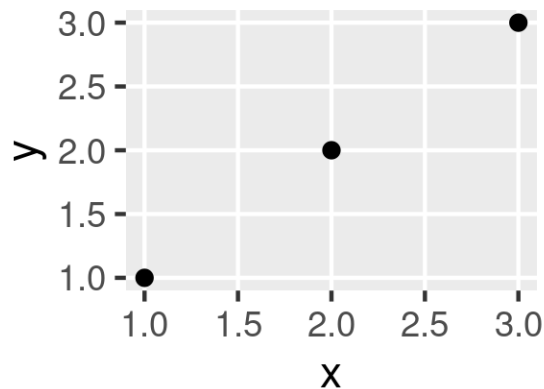
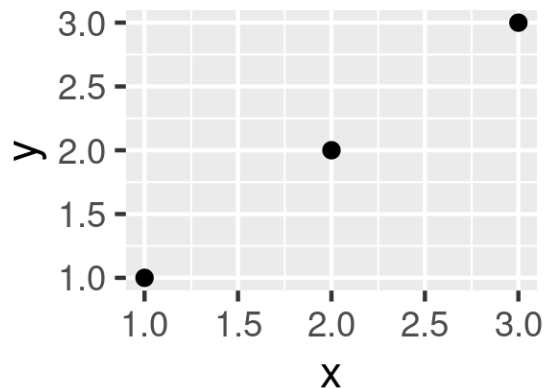


ELEMENT_BLANK

plot

`last_plot() + theme(panel.grid.minor = element_blank())`

`last_plot() + theme(panel.grid.major = element_blank())`



ELEMENTS YOU CAN MODIFY

- Plot (background, title, margin)
 - Axis (tick labels, axis titles, tick Marks, tick mark lengths)
 - Legend (key, size, height, width, margin and text alignment)
 - Panel (background, border, major and minor grids, aspect ratio)
 - Faceting (spacing between panels, backgrounds)
-
- Full List: <https://ggplot2.tidyverse.org/reference/theme.html>

APPLYING BEST PRACTICES IN R

- Modifying existing themes
- Applying a custom theme
- Alt text in markdown

MODIFYING A THEME

```
ggplot(hospital_data, aes(x = date_onset)) +  
  some_theme() +  
  theme(  
    plot.title = element_text(face = "bold", size = 12),  
    legend.background = element_rect(  
      fill = "white",  
      linewidth = 4,  
      colour = "white"  
    ),  
    legend.position = "right",  
    axis.ticks = element_line(colour = "grey70", linewidth = 0.2),  
    panel.grid.major = element_line(colour = "grey70", linewidth = 0.2),  
    panel.grid.minor = element_blank()  
  )
```

Invoking ggplot2

Applying a theme

Using the theme() function to make modifications to the theme

SAVING A THEME INTO AN OBJECT

```
mytheme <- theme(  
  plot.title = element_text(face = "bold", size = 12),  
  legend.background = element_rect(  
    fill = "white",  
    linewidth = 4,  
    colour = "white"  
  ),  
  legend.position = "right",  
  axis.ticks = element_line(colour = "grey70", linewidth = 0.2),  
  panel.grid.major = element_line(colour = "grey70", linewidth = 0.2),  
  panel.grid.minor = element_blank()  
)
```

SAVING LABELS INTO AN OBJECT

```
my_labels <- list(ylab("my y axis"),  
  xlab("my x axis"),  
  ggtitle("my title")  
)
```

APPLYING THEME AND LABEL OBJECTS

```
my_theme <- theme(...)
```

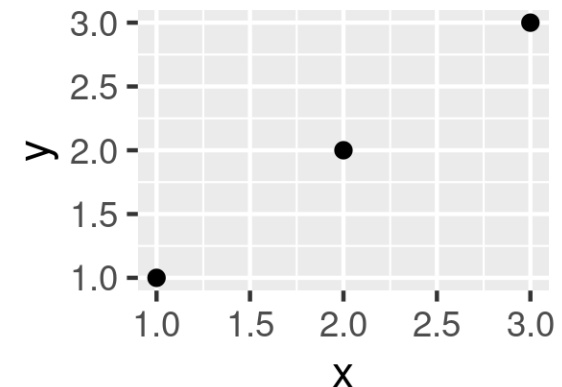
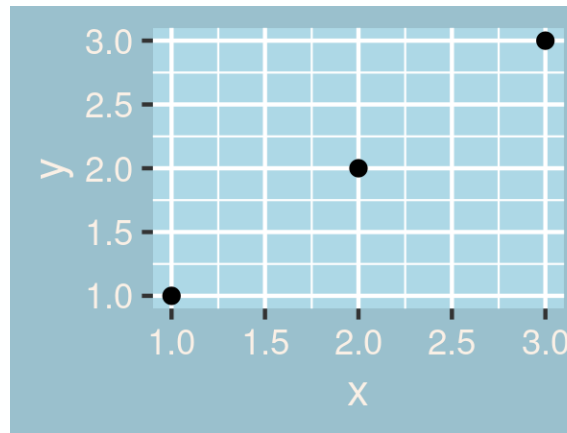
```
my_labels <- list(...)
```

```
ggplot(hospital_data, aes(x = date_onset)) +  
  my_theme + my_labels
```

UPDATING AND REVERTING THEME OBJECTS

```
old_theme <- theme_update(  
  plot.background = element_rect(fill = "lightblue3", colour = NA),  
  panel.background = element_rect(fill = "lightblue", colour = NA),  
  axis.text = element_text(colour = "linen"),  
  axis.title = element_text(colour = "linen")  
)
```

```
plot  
theme_set(old_theme)  
plot
```



ALT TEXT IN MARKDOWN

```
17 ▾ ```{r fig.cap="Some Health Data", fig.alt = "A scatterplot chart showing some  
relationship between an intervention and an outcome"}  
18 ggplot(data = health_data, aes(x = intervention,  
19   y = outcome,  
20   color = comparator)) +  
21 geom_point(aes(shape = comparator), alpha = 0.8) +  
22 scale_color_manual(values = c("darkorange", "purple", "cyan4"))  
23 ▲ ```
```

```
17 ▾ ```{r, fig.alt = "Crochet (not knitting!) needle with colorful yarn",  
out.width="25%"}  
18 knitr::include_graphics("thumbnail.jpg")  
19 ▲ ```
```

A background image showing two women in an office environment. One woman is pointing at a laptop screen while the other looks on. The image is darkened with a blue overlay. A large, bright blue curved shape is in the bottom right corner.

“WORST” PRACTICES IN DATA VISUALIZATION

Part 1

SMALL GROUP ACTIVITY

- Combine your knowledge of data visualization best practices with that of the `theme()` function!
- Groups will be provided with a dataset and base code to which they will apply theming elements that fly in the face of common decency and best practices
- You will have 20 minutes to break the rules and compete with your colleagues to create a figure that is even **WORSE** than the example provided!

A background image showing two women sitting at a table, looking at a tablet together. The woman on the right is smiling. The image is overlaid with a dark blue semi-transparent filter. A bright blue curved shape is at the bottom right.

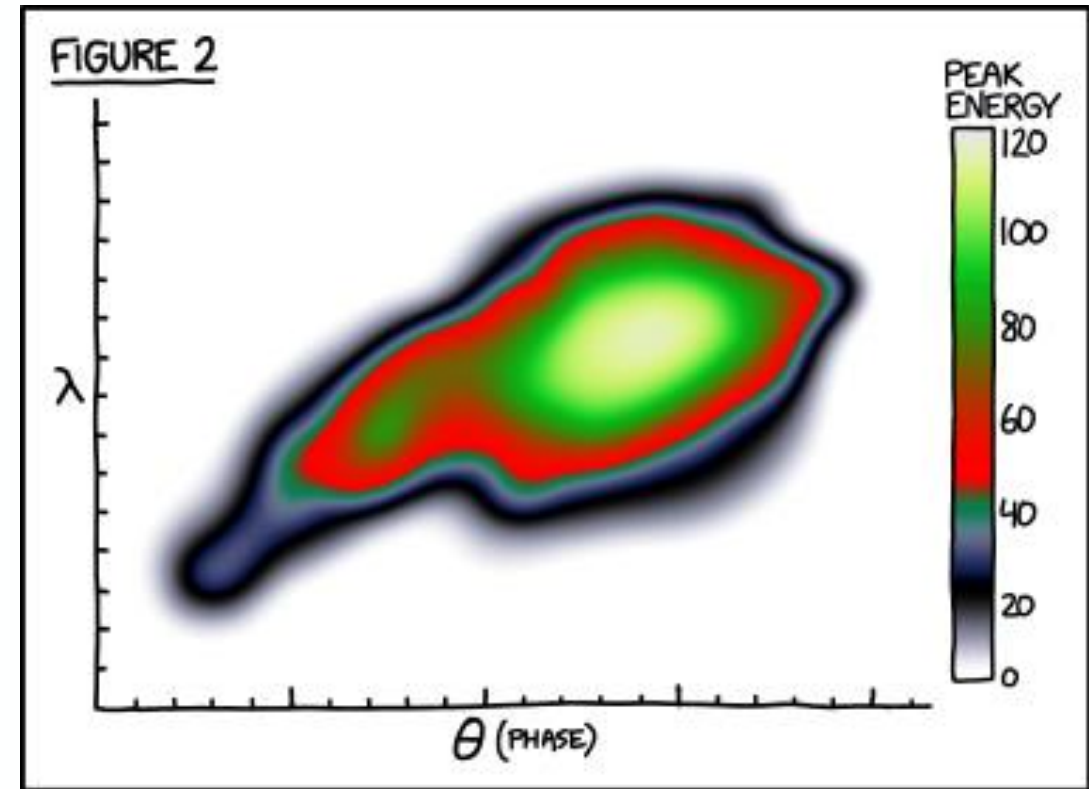
“RULES TO VIZ BY”

SMALL GROUP ACTIVITY: RULES TO 'VIZ' BY

- Let's all grow from what we witnessed in the last activity and commit to never visit that terrible place again!
- Based on what was presented and learned from the "worst" practices in data visualization, develop a list of rules to live by for data visualization and its applications in R
- Using the provided jamboard:
 - Add as many rules you can think of!
 - Select your top three 'rules' before coming back so we can compile them into a handout!

BREAK!

Please return to the
virtual classroom by:
3pm ET



EVERY YEAR, DISGRUNTLED SCIENTISTS COMPETE
FOR THE RAINBOW AWARD FOR WORST COLOR SCALE.

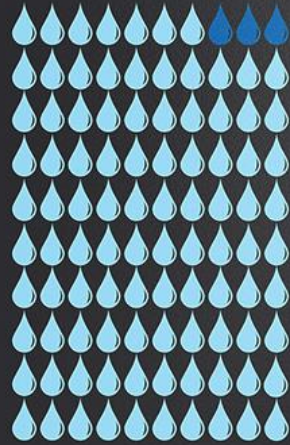
Source: <https://xkcd.com/2537/>

A background image showing two women, one with long dark hair and one with curly dark hair, both smiling and looking at a laptop screen. The image is overlaid with a dark blue semi-transparent filter. A bright blue curved shape is at the bottom right.

“WORST” PRACTICES IN DATA VISUALIZATION

Part 2

Total World Water



GLOBAL AVERAGE WATER
FOOTPRINT IS AROUND
1,240
M³ PER YEAR PER PERSON

70% OF EXISTING FRESH
WATER IS WITHDRAWN
FOR IRRIGATION IN
AGRICULTURE

10 gallons
ON AVERAGE
PER DAY OF YOUR
WATER FOOTPRINT
IS LOST TO LEAKS
(10% OF YOUR
INDOOR USE)



• RIVERS AND LAKES

• GROUND WATER

• ICE CAPS

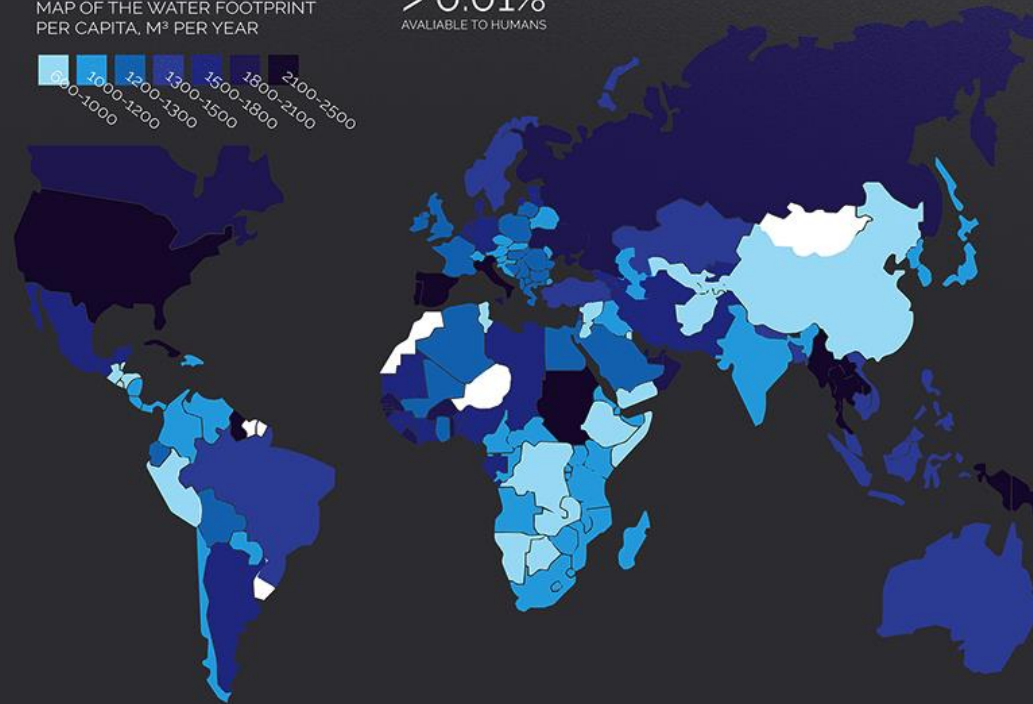


1 in 8
people
DO NOT
HAVE ACCESS
TO CLEAN WATER

MAP OF THE WATER FOOTPRINT
PER CAPITA, M³ PER YEAR

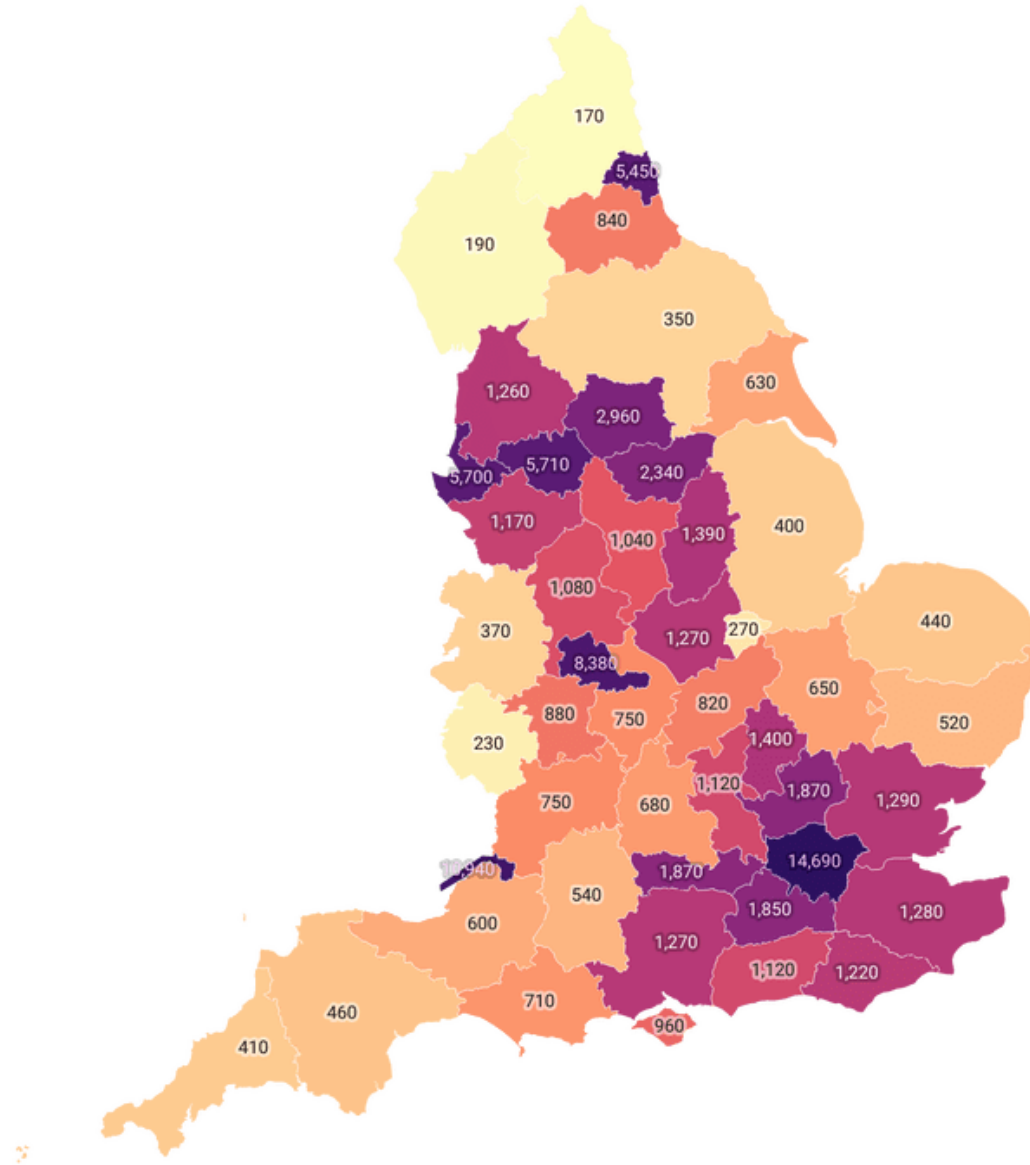


> 0.01%
AVAILABLE TO HUMANS



Population density of each county in England

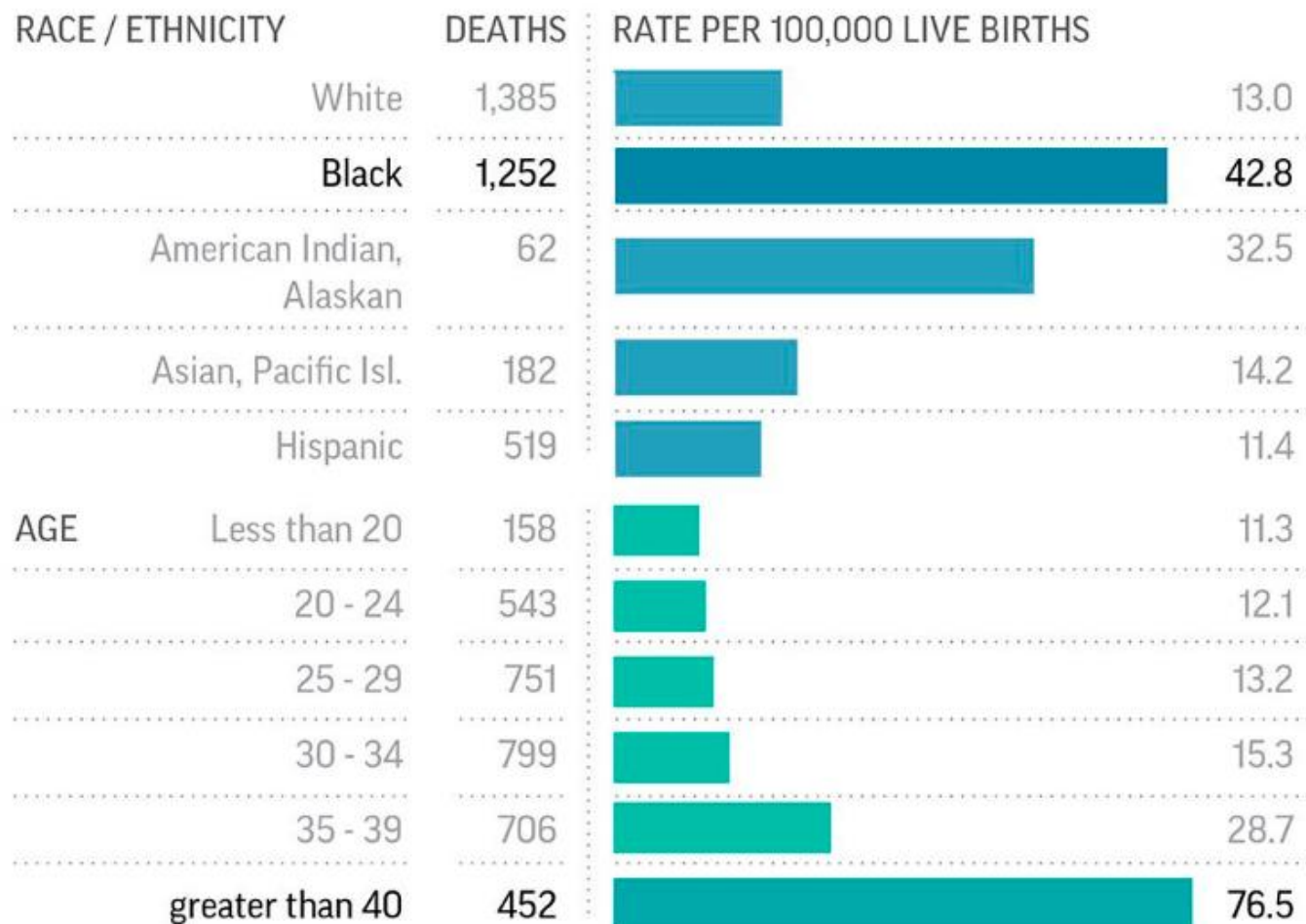
In Square Miles



Created with Datawrapper

Pregnancy deaths rare but higher in some groups

A new federal report finds that pregnancy-related deaths are rising in the U.S., especially among black women.

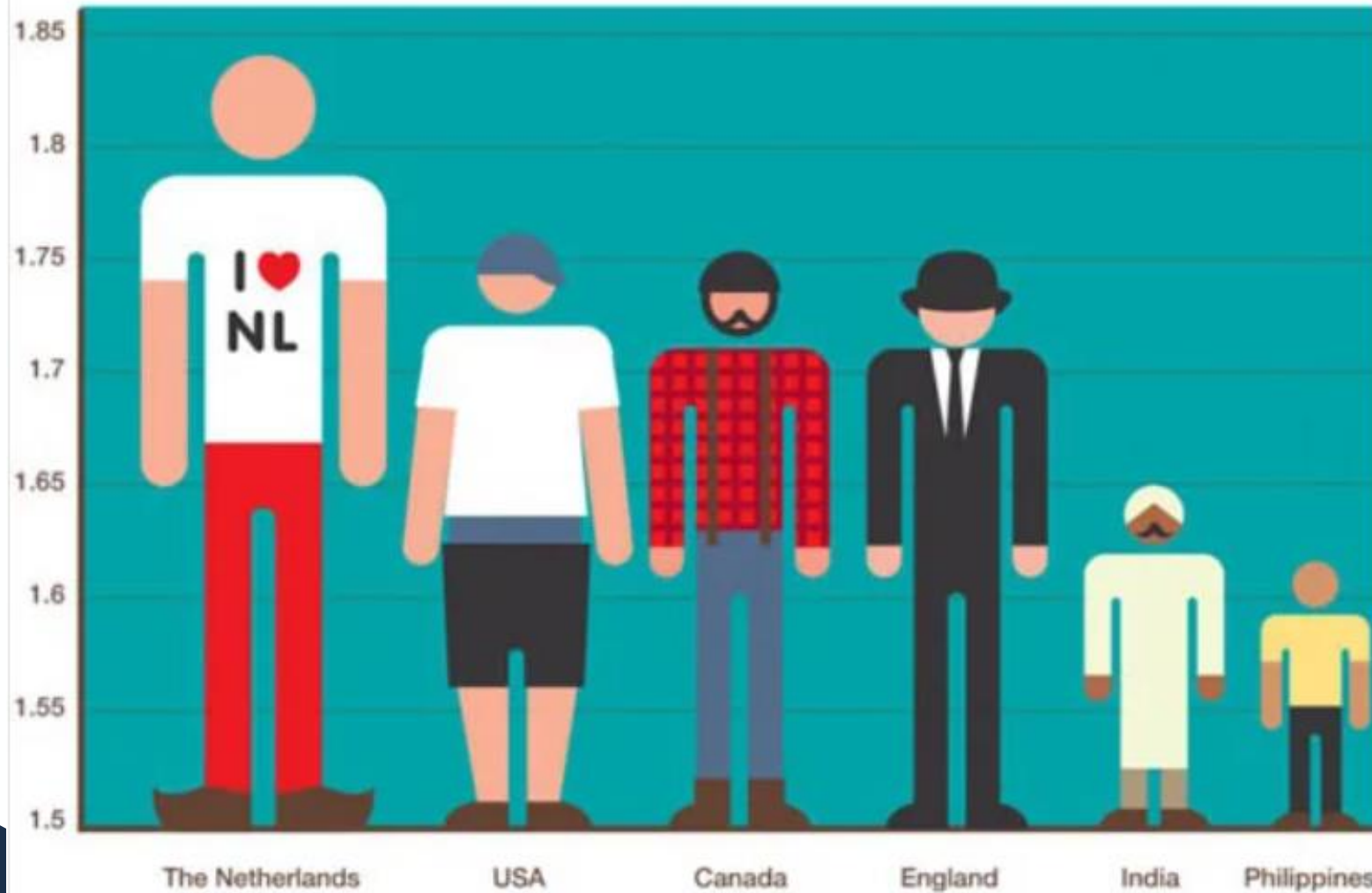


SOURCE: Centers for Disease Control and Prevention, 2011-2015 data

AP

LOOKING DOWN ON THE REST OF THE WORLD

(Average male height in m)



DO NO HARM

Picture this: Doing Good Data Means Doing No Harm

EQUITY, DIVERSITY & INCLUSIVENESS

***“If I were one of the data points on this visualization,
would I feel offended?”***

- Journalist Kim Bui in the Do No Harm Guide

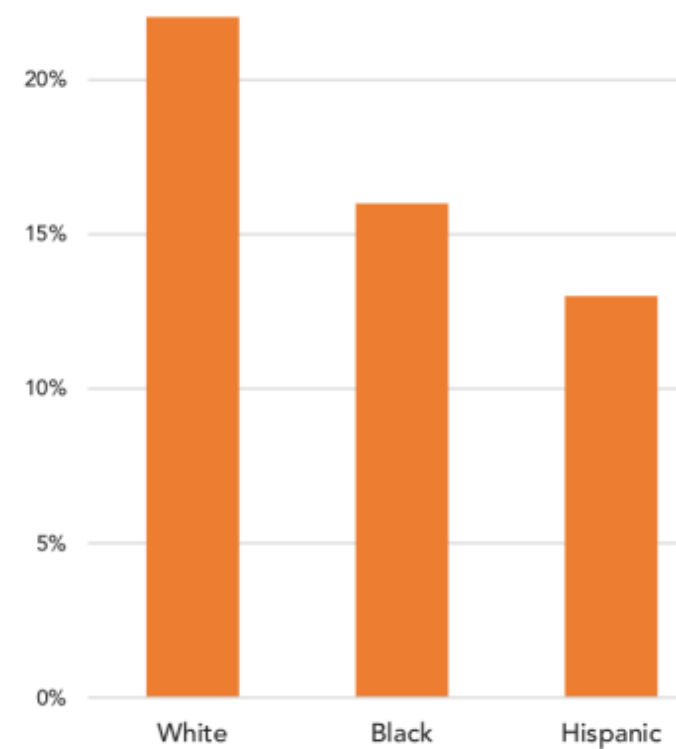
EDI RECOMMENDATIONS (1 OF 6)

- Critically examine your data
 - Who (inclusion/exclusion), how, why, benefits and harms
- Use people-first language
 - Ex: 'people with disabilities' or 'a person with asthma'
- Label people, not skin colour
 - Language continues to evolve

EDI RECOMMENDATIONS (2 OF 6)

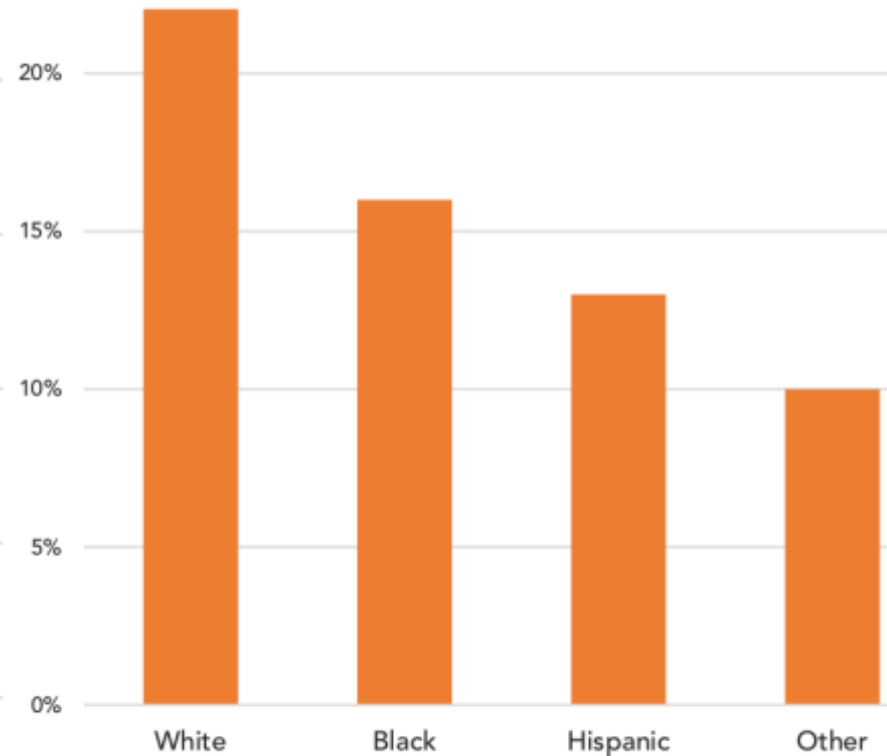
Mental Health in Jail

Rate of mental health diagnosis of inmates



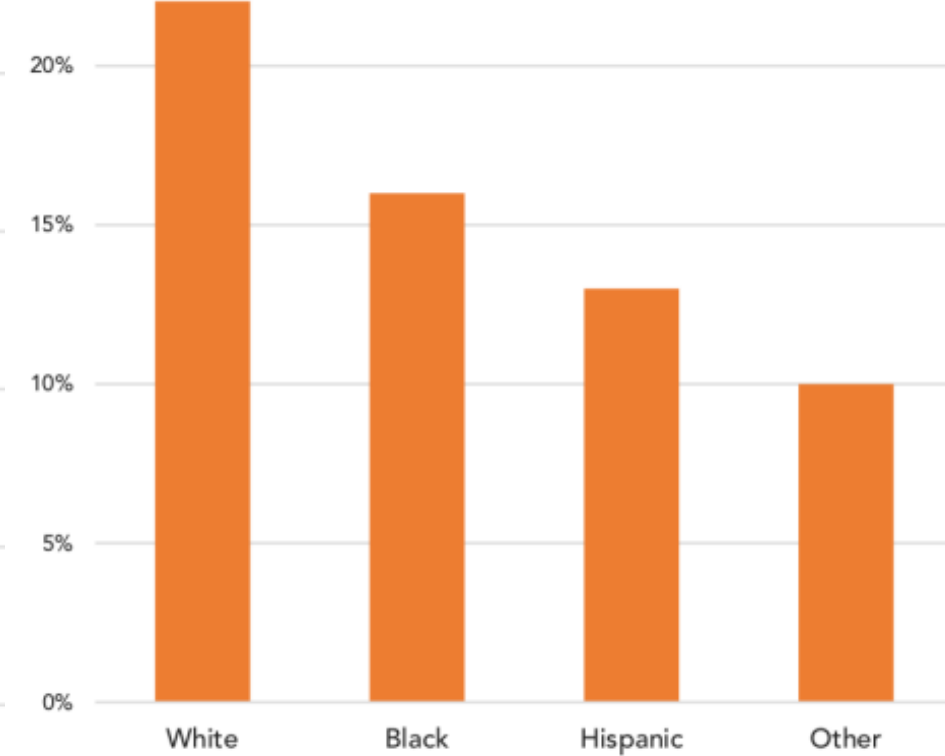
Racism in Jail

People of color less likely to get mental health diagnosis



Racism in Jail

White people get more mental health diagnoses



EDI RECOMMENDATIONS (3 OF 6)

- Orders labels purposefully
 - Study focus, specific story, quantitative relationship (magnitude), alphabetical, etc.
- Consider missing groups
 - Highlight how the data are not inclusive or representative
 - Think of other options for 'other'

25. Is this person:

Mark "x" more than one circle or specify, if applicable.

- White
- South Asian (e.g., East Indian, Pakistani, Sri Lankan)
- Chinese
- Black
- Filipino
- Arab
- Latin American
- Southeast Asian (e.g., Vietnamese, Cambodian, Laotian, Thai)
- West Asian (e.g., Iranian, Afghan)
- Korean
- Japanese
- Other group — specify:

EDI RECOMMENDATIONS

(4 OF 6)

<p>Race</p> <p><i>In our society, people are often described by their race or racial background. These are not based in science, but our race may influence the way we are treated by individuals and institutions, and this may affect our health. Which category(ies) best describes you? Select all that apply</i></p>	<ul style="list-style-type: none"><input type="checkbox"/> Black (African, African Canadian, Afro-Caribbean descent)<input type="checkbox"/> East Asian (Chinese, Japanese, Korean, Taiwanese descent)<input type="checkbox"/> Indigenous (First Nations, Inuk/Inuit, Métis, Other, please specify below)<input type="checkbox"/> Latin American (Hispanic or Latin American descent)<input type="checkbox"/> Middle Eastern (Arab, Persian, West Asian descent (e.g., Afghan, Egyptian, Iranian, Kurdish, Lebanese, Turkish)<input type="checkbox"/> South Asian (South Asian descent, e.g., Bangladeshi, Indian, Indo-Caribbean, Pakistani, Sri Lankan)<input type="checkbox"/> Southeast Asian (Southeast Asian descent, <u>e.g.</u> Cambodian, Filipino, Indonesian, Thai, Vietnamese, Laotian, Malaysian)<input type="checkbox"/> White (European descent)<input type="checkbox"/> Another race category, specify:<input type="checkbox"/> Unknown<input type="checkbox"/> Declined to answer
--	--

EDI RECOMMENDATIONS (5 OF 6)

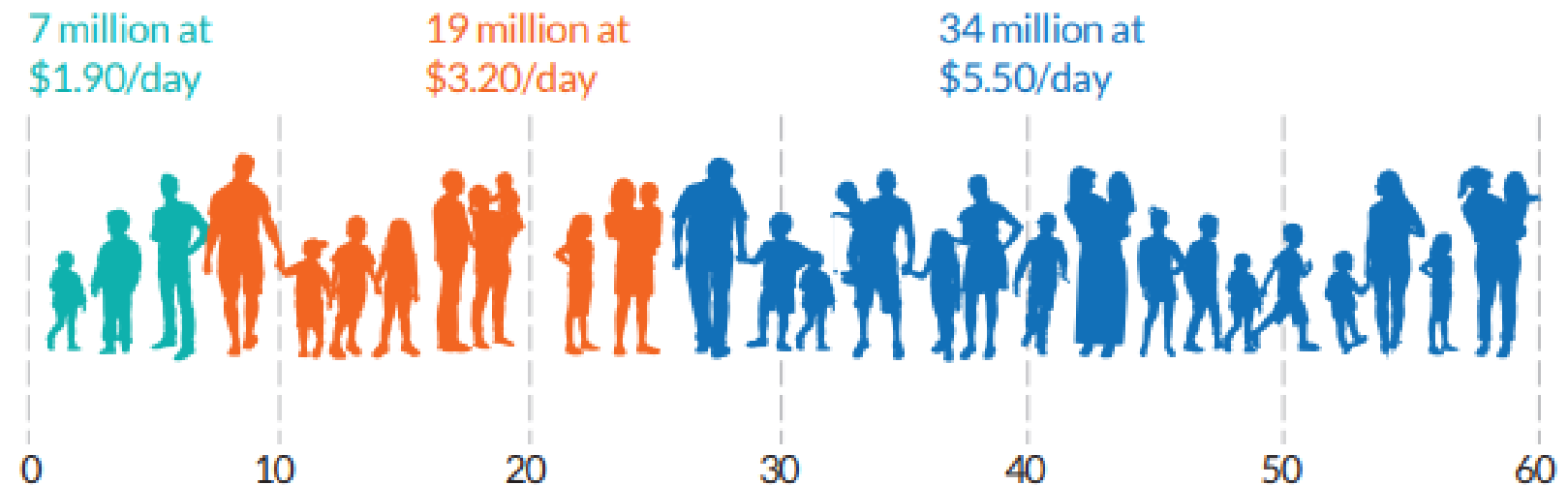
- Carefully consider colours

Millions of people in poverty



- Consider icons and shapes

Millions of people in poverty



Source: Recreated based on Schwabish (2021).

EDI RECOMMENDATIONS (6 OF 6)

- Communicate with people and communities of focus
 - Collaboration
- Reflect lived experiences
 - Consider what your work may be missing
- Consider the needs of your audiences
 - Useful format
 - Appropriate language
 - Translation

DISCUSSION: EXAMPLES IN THE WORKPLACE

- Do you have any examples where data visualizations have gone wrong?

PROTECTING PRIVACY

Measures to protect against re-identification:

- Suppress, mask or redact direct identifiers
- Assess the risk of re-identification through indirect identifiers
- **When releasing data in tables, determine the appropriate minimum cell size**
- Modify the data to mitigate the risk of re-identification
- Document the process
- Plan for regular, ongoing and periodic assessment of re-identification risk

DISCUSSION: MINIMUM CELL SIZE

- What are some considerations of determining minimum cell size?
- What can we do if we have small cell sizes?

LARGE GROUP ACTIVITY

- Visit the Jamboard provided and list as many considerations as you can for ways that data visualization can contribute to harm among vulnerable populations related to either **person**, **place**, or **time**. Feel free to come off of mute to describe any examples you can think of or write it directly on the Jamboard.

Person

- E.g., How are people in the visualization represented? Is there potential for offense? Stigmatization?

- **Place**

- E.g., How have geolocations been aggregated? Can represented areas be used to draw hurtful conclusions about certain populations?

Time

- E.g., How has calendar time or dates been represented? Does it align with cultural events?

A background image showing two women sitting at a table in a library or study. The woman on the right is smiling and pointing at a tablet. The woman on the left is looking at the tablet. There are bookshelves in the background. The image is overlaid with a dark blue semi-transparent layer.

DEMO

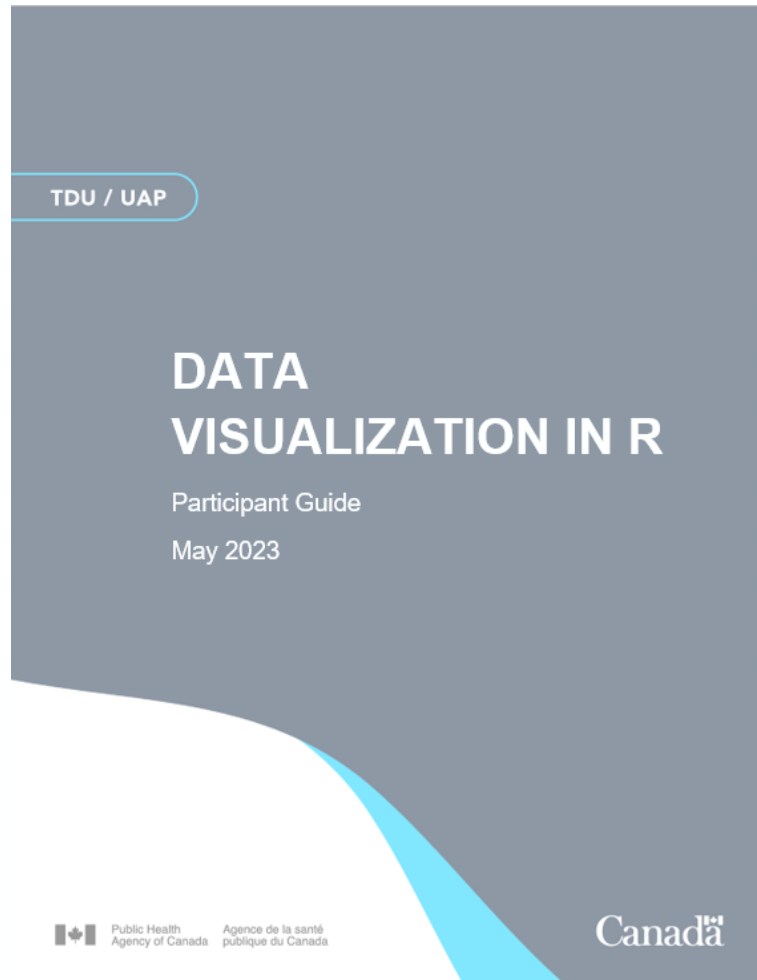
A background image showing two women sitting at a table in a library or study. One woman is pointing at a tablet held by the other, who is smiling. Bookshelves are visible in the background. The image is overlaid with a dark blue semi-transparent layer.

WRAP-UP

SUMMARY AND WRAP-UP

- By the end of this course, participants will be able to:
 - ✓ Discuss differences between base R and the grammar of graphics (ggplot) coding styles for data visualization.
 - ✓ Apply knowledge of R-coding to automate common graphics used in public health.
 - ✓ Connect elements of effective graphic design to R coding practices in data visualization.
 - ✓ Discuss considerations for data visualization to avoid potential harms to small or vulnerable communities.

SUMMARY AND WRAP-UP



- Independent activity 2
- Creating graphics that:
 - Adhere to data visualization best practices
 - Take into account harms to a vulnerable community

END OF DAY CHECK-IN

With the jamboard link provided, use the post-it note feature to share either:

- One word you would use to describe the day

OR

- One thing you would like to have learned in days 1 and 2 of the course that was not covered