


ENVIRONMENTAL DATA ANALYTICS: M5 – DATA VISUALIZATION

Spring 2023

Nicholas School of the Environment - Duke University

Catch up

- **Debugging**:
  - Goal: find where the error might occur…
  - Start simple, add complexity in increments…
  - Check outputs for logical consistency…

- **Knit issues**:

```
Calls: <Anonymous> ... eval_with_user_handlers -> eval -> eval -> install.packages -> contrib.url
```

  - Check paths (knit directory = project working directory)
  - Restart R (clear environment) and run entire Rmd file…

M5.1- Data Visualization

- Approaches to visualizations (link)
- The ggplot2 package
- ggplot structure: layers = geoms
- Aesthetics, axes, colors, shapes, facets, axis limits, reference lines
- Plot types…

M5.2 – Formatting plots

- Themes
- Custom layers
- Color palettes
- Cowplots package
- Saving plots

M4.3 – Data Visualization III (lab)

Expressions

- geom_text()
  - The paste and paste0 commands
  - MathJax

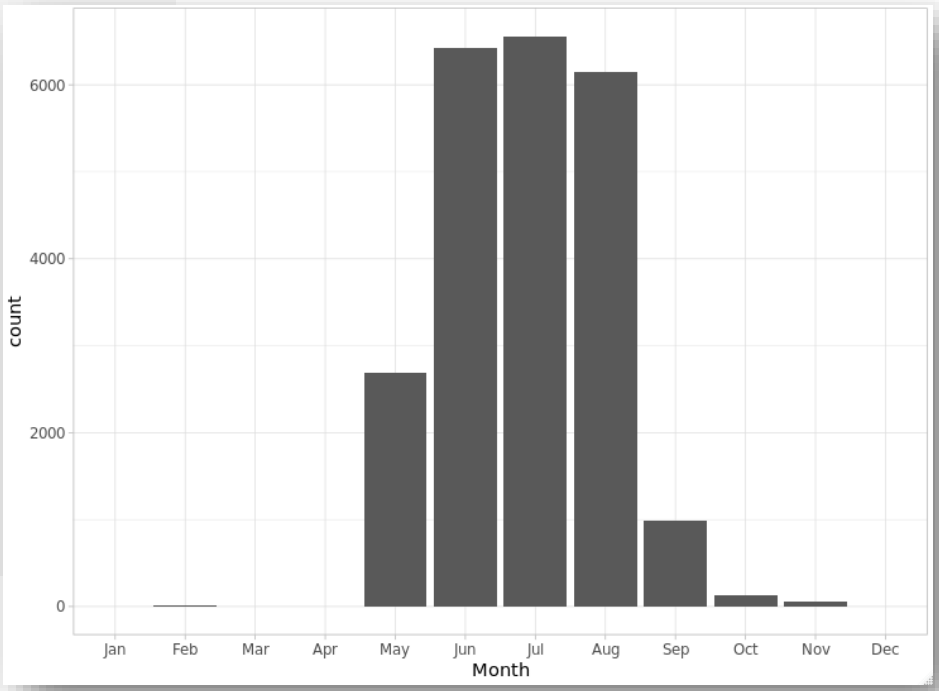
A note on **factors**…

- *Factors..*
  - …are useful for analyzing/visualizing categorical data
  - …have *levels*
  - …can have *labels* too

- *Plot the number of lake measurements by month…*
  - What kind of variable is `month` in the dataframe?
  - How many unique values in this column?
  - Why might this pose a problem?
  - What can we do?

A note on factors… *Solution*

```
#Tidy up the code
the_plot <- PeterPaul.chem.nutrients %>%
  ggplot(
    aes(x=factor(
      month,
      levels=1:12,
      labels=month.abb)
      )
    ) +
geom_bar() +
scale_x_discrete(
  name="Month",
  drop=FALSE
  )
#Show the plot, in the light theme
the_plot + theme_light()
```



More on *themes…*

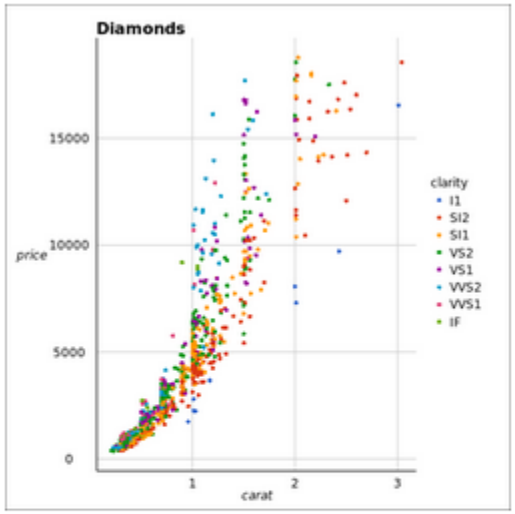
Themes control the following elements

- **Plot background**:
  The background color or fill pattern of the plot area.
- **Plot title**:
  The size, font, and position of the plot title.
- **Axis labels**:
  The font, size, and position of the x-axis and y-axis labels.
- **Axis ticks and grid lines**:
  Color, size, & position of the tick marks & grid lines on the axes.
- **Legend**:
- The font, size, and position of the legend.

ggthemes()

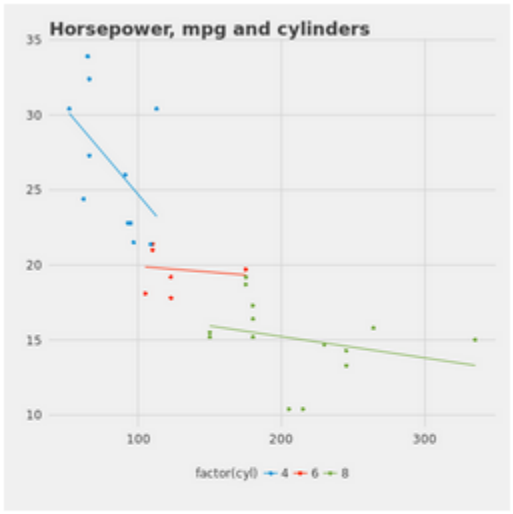
Adds custom themes and scales

- Link to examples



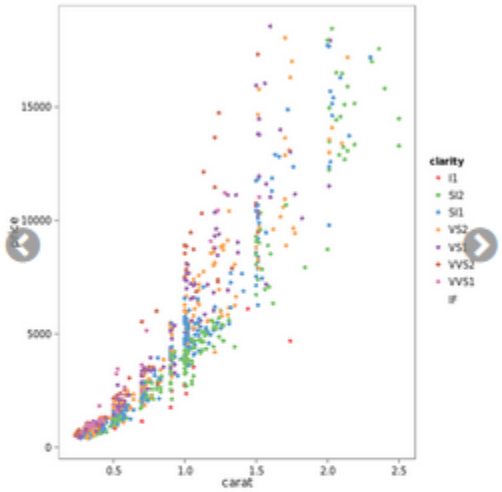
theme_gdocs

Theme with Google Docs Chart defaults



theme_fivethirtyeight

Theme inspired by fivethirtyeight.com plots

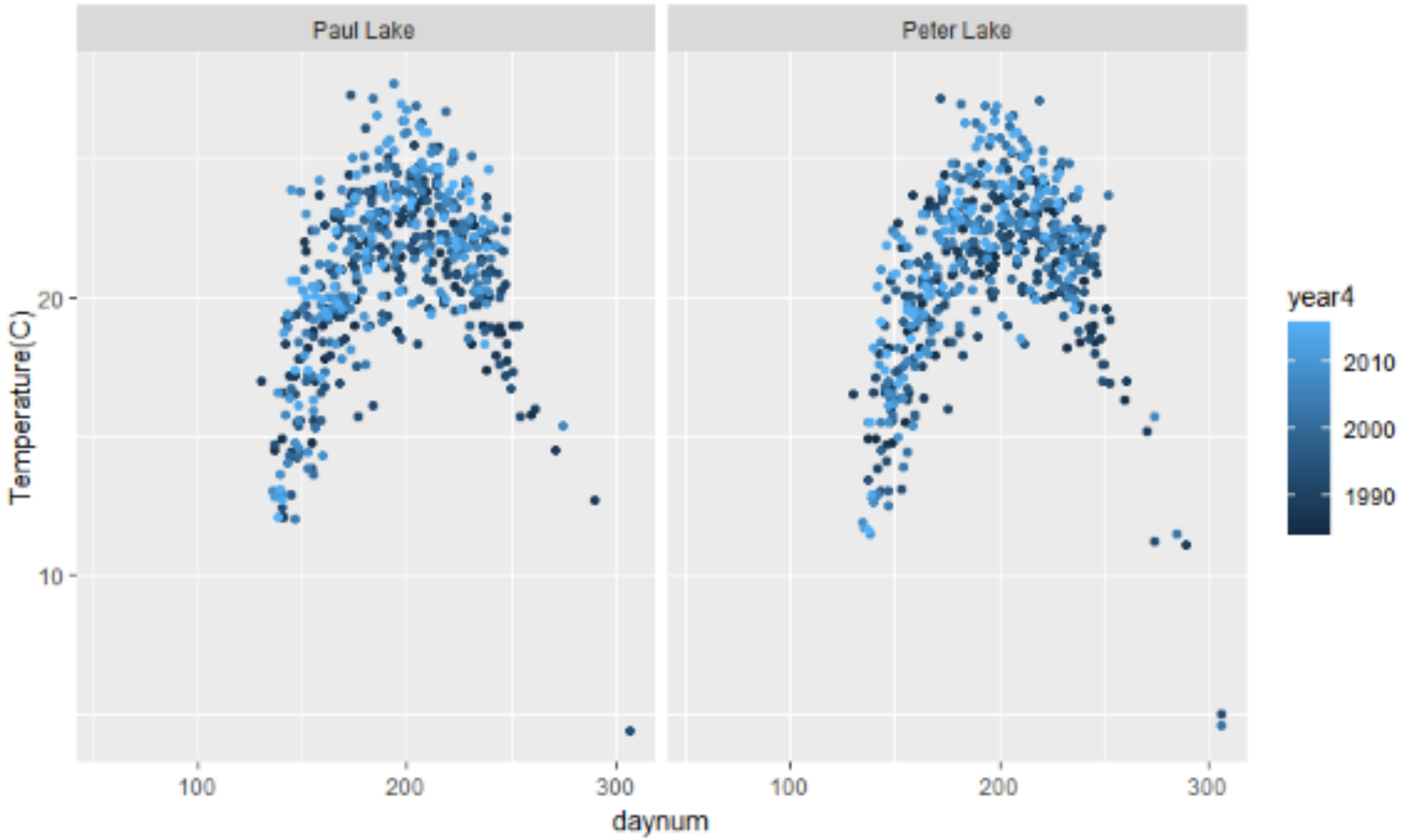


theme_few

Theme based on Few's "Practical Rules for Using Color in Charts"

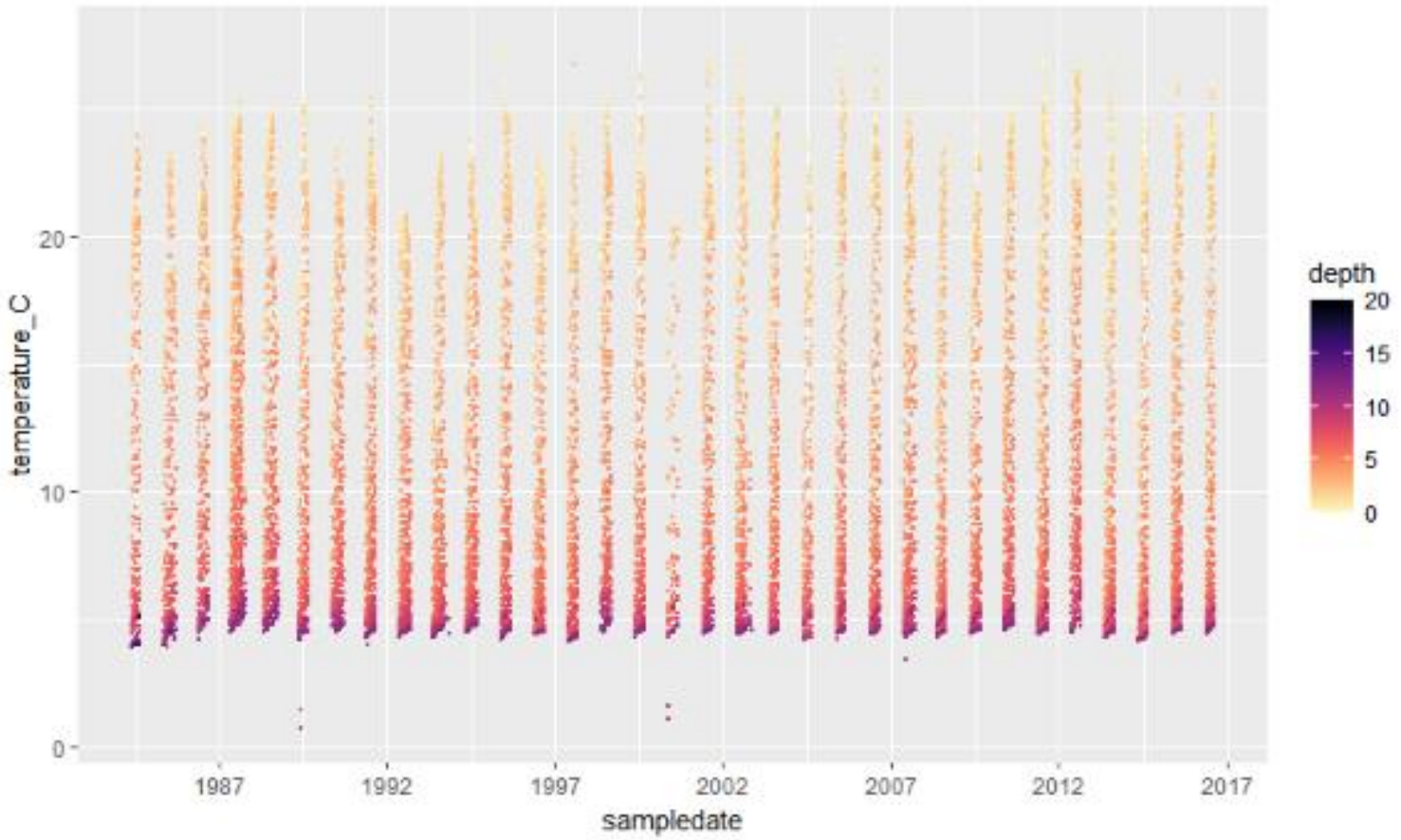
Exercise 1

```
# 1.
# Plot surface temperatures by day of  year.
# Color your points by year, and facet by lake in two rows.
# Change the ylab name
```



Exercise 2

```
#2.
# Plot temperature by date. Color your points by depth.
# Change the size of your point to 0.5
# Change the color palette to magma and play with direction (+- 1), which one makes more sense?
# Change x axis to include marker/labels every 5 years
```

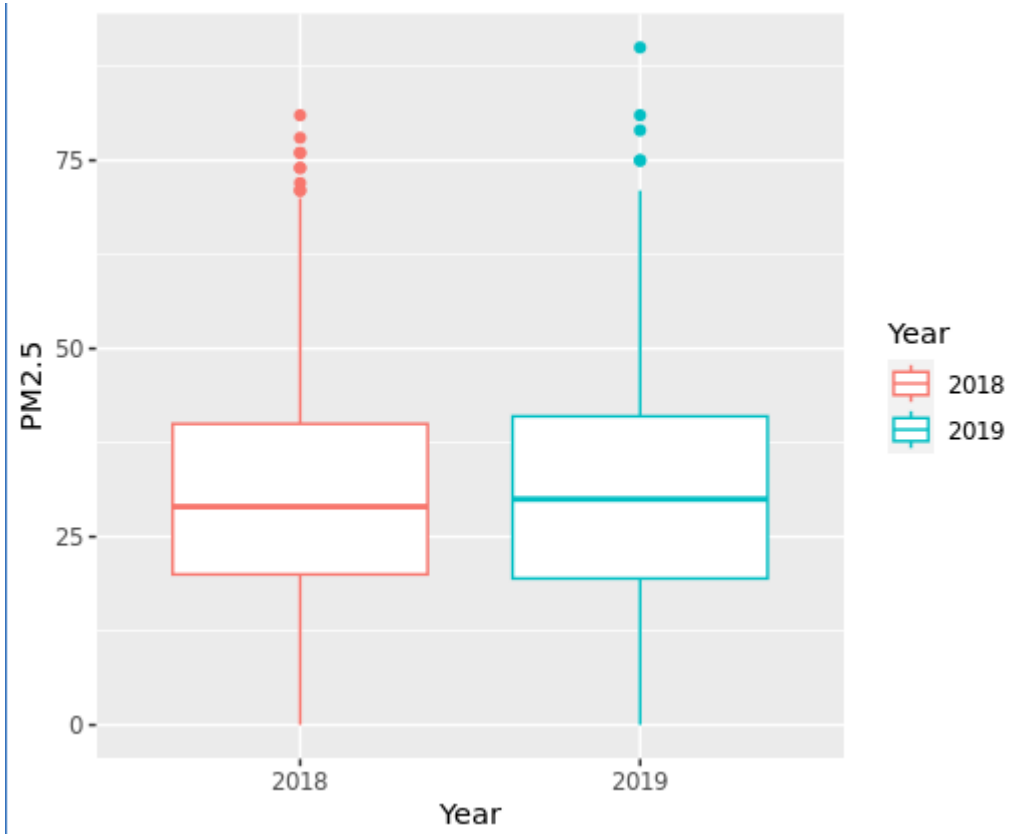


Exercise 3

```
# 3.
# Plot AQI values for ozone by PM2.5, colored by latitude
# Make the points 50 % transparent
# Add a line of best fit for the linear regression of these variables.
```

Exercise 4

```
# 4.
# Create several types of plots depicting PM2.5, divided by year.
# Choose which plot displays the data best and justify your choice.
```



Viz challenge?