With your group, create a slide with the two parts below,

- Solution: Provide a template answer to the problem. Highlight any aspect that you would like to discuss in the follow-up period.
- Brainstorm: What do you think would be a related question given this topic?

Also submit your team’s slides to gradescope: https://go.wisc.edu/7087tb

Practice Exam: https://go.wisc.edu/77jxs6

- Q1 - Groups 1, 2
- Q2 - Groups 2, 3
- Q3 - Groups 3, 4
- Q3 - Groups 4, 5
- Q4 - Groups 6, 7
- Q4 - Groups 7, 8
- Q5a - Groups 8, 9
- Q5b - Groups 9, 10
- Q6a - Groups 10, 11
- Q6b - Groups 11, 12
- Q7 - Groups 12, 13
- Q8a - Groups 13, 14
- Q8b - Group 14

Question 1 - Group 1

Solution:

A and C are true.

We think A is true because if you were grouping by date, each graph would be a different date

B is not true because it would be facet_grid to be vertical

C is true because “+” concatenates horizontally, “/” concatenates vertically

We think D is not true because you only encode x in density plots

Brainstorm: A related question given this topic would be True or False: If p1, p2, and p3 are ggplot objects, the code `p1 / p2 / p3` concatenates three figures vertically

Question 1&2 - Group 2

Question 1:

a & c are true.

Question 2:

Small Multiples(D): The use of the same visual encoding to several partitions of the dataset, shown together across adjacent panels.

Details-on-demand(C): A detail view that appears when one or more items are selected in the main view

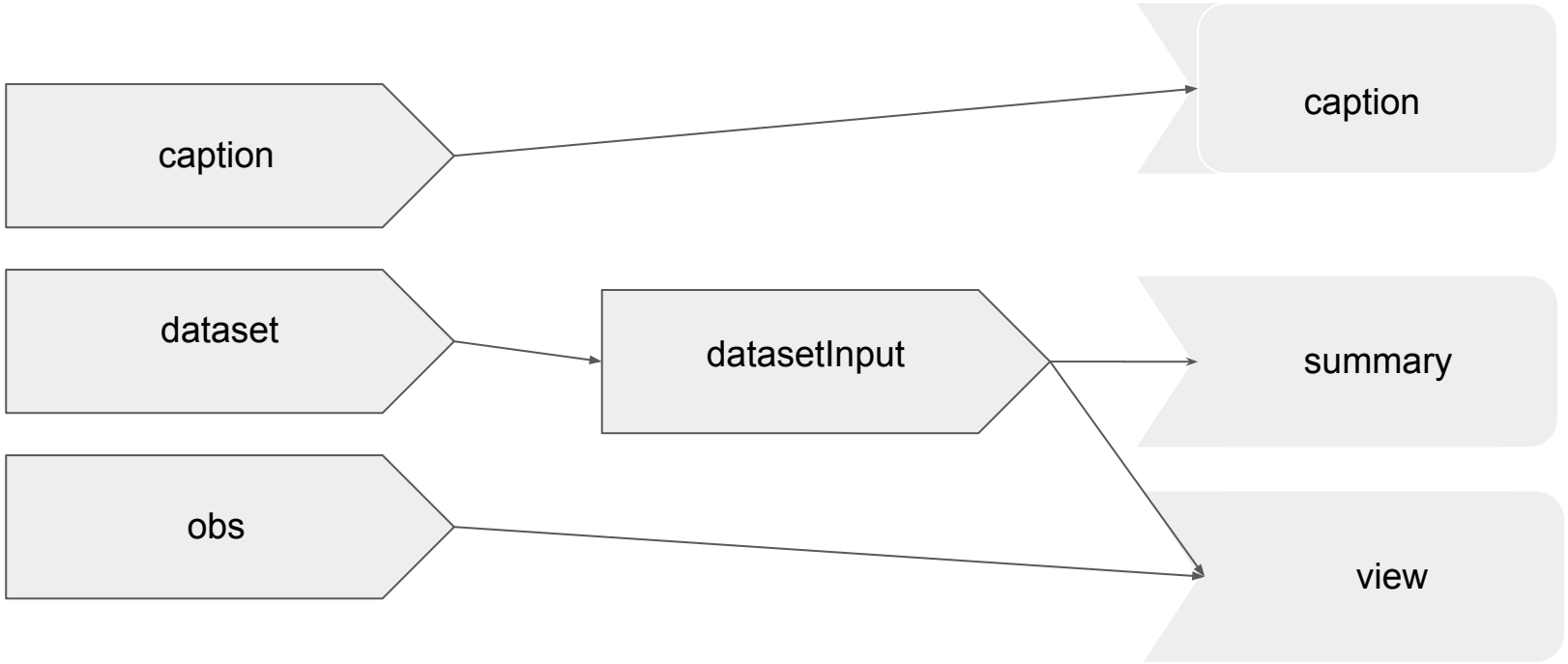
Multiform display(B): The use of different visual encodings to represent the same data, shown together across adjacent panels.

Linked Brushing(E): A practice where items interactively selected in one overview are highlighted in all other views.

Further discussion: Write out the definitions of all six terms in the options, and **highlight their differences**, then **plot** to help understand

Similar question: Given several graphs and several definitions to match with one another(aiming to determine which is geom_ridge_density plot and which is geom_ridgeline plot)

Question 7 - Group 13



Question 8A - Group 13

```
birds = birds %>%
        mutate(Order = tolower(Order))
df = birds%>%
        group_by(Order)%>%
        summarize(meanAsymmetry = mean(Asymmetry))
```

Related Q : What is the mean asymmetry across each family of birds, in order of asymmetry?

Question 4 - Group 6

```
m %>%

  pivot_longer(-country, names_to = "medal", values_to = "value")
```

arrange the data frame

- Brainstorm : transform longer dataset to wider & wider to longer
- Relevant data preprocessing issues

Q3 - Groups 4, 5

- Inconsistent visual encoding (i.e. making m and f different colors across plots). This can be mitigated by considering how variables will be represented and making things like size, shape, axis, and color coding consistent across plots.
- Inconsistent Alignment- Poor alignment of plots can both take away from aesthetic quality and lead to misinterpretation. Making sure that plots are in a consistent square or rectangle and making sure axes are aligned to emphasize what patterns the viewer should notice.

Question 4, 5a - Group 8

4. pivot_longer(medals, c(“gold”, “silver”), names_to = “medal”, values_to = “value”)

Related: pivot_wider question for the reverse, question about tidy data

5a. Moved legend to top of plot, added title and modified axes labels, changed axes tick labels, changed color palette, added trend line, changed background (color, minor axes are removed, changed color of major axes), added labels to points

Related: Write code for the modification described, describe how the modifications aid the visualization (may be too subjective)

Question 5b & 6a - Group #10

5b. -xlab(“Population(logscale)”)

-geom_smooth(se = FALSE, method = “lm”)

-ggplot(df, aes(population, total, label = state_abbreviation))+

geom_point()

Question 3 - Group 4

Discuss two common pitfalls in compound figures. What can be done to avoid them?

1. Inconsistent Encodings for Shared Variables
    a. You should instead aim to encode the same color scheme across all panels so all the figures use the same colors for the particular variables.
2. Obtrusive Annotation
    a. You should have visible labels – not too large, consistent fonts across figures, logical ordered
    b. Should be thought of like page numbers; they are useful to make reference to, but don’t necessarily have to be read

Question 4 - Group 7

```
medal %>%
  pivot_longer(
            cols=gold:silver,
            names_to = “medal”,
            values_to = “value”)
```

Or

```
medal %>%
pivot_longer(c(“gold”, “silver”), names_to = “medal”, values_to = “value”)
```

Question 2 & 3 - Group 3

Q2: A. Ridgeline plot B. Multiform display C. Details-on-demand D. Small multiples E. Linked brushing F. Tidy data

• **D** The use of the same visual encoding to several partitions of the dataset, shown together across adjacent panels.

• **C:** A detail view that appears when one or more items are selected in the main view.

• **B:** The use of different visual encodings to represent the same data, shown together across adjacent panels.

• **E:** A practice where items interactively selected in one overview are highlighted in all other views.

Brainstorm: Given a figure/graph and asked to identify what terms best describe it.

Q3: Two common pitfalls of compound figures. How to avoid them?

- The plots do not align by the edge, and the color encoding is different among in each plot which makes harder for readers to analyze the plots cohesively.
- Patchwork syntax like +, / will solve the alignment and color legend problems.
    - Combine legends using plot_layout(guides = “collect”)

Brainstorm:

Questions 6a and 6b - Group 11

- 6a: geom_ridgeline() takes a set of heights and plots them directly, filling in the area underneath. geom_density_ridges() works similarly, but instead of directly plotting heights from a table it first takes a dataset and creates a list of heights base on the density of certain values.
- 6b: geom_ridgeline() could be used to track a value over time - for example, the population of fish in a river. geom_density_ridges() could be used to display the distribution of scores on a test.
- Brainstorming: how could you modify a geom_ridgeline() plot to encode an additional feature?

Question 5a, 5b - Group 9

5a.

- Added title, x-axis, and y-axis labels
- Changed display of x-axis ticks (from scientific notation to “n million” for example)
- Added point annotations

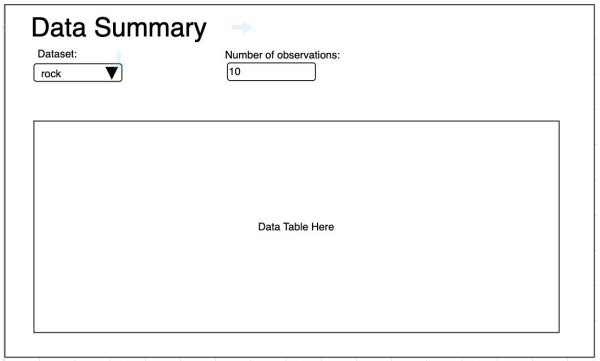
5b.

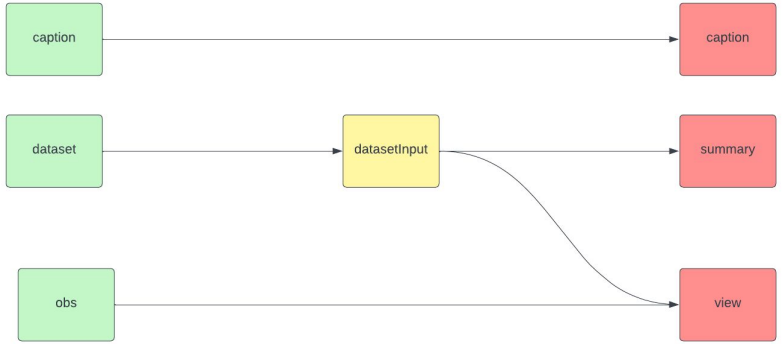
- labs(x=”Population (log scale)”, y=”Total number of murders (log scale)”, color = “Region”, title=”US Gun Murders in 2010”)
- scale_x_log10(labels = unit_format(unit = "million", scale = 1e-6))
- geom_text_repel(aes(label = abb), segment.size = 0.2)

Question 6b and 7 - Group 12

6b. Geom_ridgeline has direct values associated with the height line where geom_ridge_density provides the density of the variables values (like a histogram). You could use geom_ridgeline when looking at specific heights over time and you could use geom_ridge density when looking at a relative distribution for a collection of variables.

7.





A related question could be asking you to write the R code for a Shiny App given a reactive graph.

Question 8a - Group 14

Provide code to convert the ‘Order’ variable to lower case and to compute the mean ‘Asymmetry’ within each order:

```
birds %>%
  mutate(Order = tolower(Order)) %>%
  group_by(Order) %>%
  summarize(mean_asymmetry = mean(Asymmetry))
```

Question 8b - Group 14

Using the transformed data from part a,

```
ggplot(birds, aes(x = reorder(Order, mean_asymmetry), y = mean_asymmetry)) +
  geom_col(width = 1) +
  coord_flip() +
  xlab("Order") +
  ylab("Mean Asymmetry") +
  scale_y_continuous(expand = c(0,0)) +
  theme(panel.grid = element_blank(),
        axis.ticks = element_blank())
```

Question # - Group

Question # - Group

Question # - Group