Session 11: EDA — Covariation — Pen-and-Paper Pair Exercise

PSY 410 | Data Science for Psychology

Name: _____ **Date:** _____

No laptop today? No problem. This handout lets you practice the same skills on paper. Work with a partner who has a laptop and compare your work at the end.

The data: `therapy_data`

This dataset has post-treatment depression scores (BDI-II) for 150 participants across three conditions. Here are 12 representative rows:

condition	depression_post
Control	20.3
Control	15.8
Control	22.1
Control	18.6
CBT	11.4
CBT	14.2
CBT	9.8
CBT	13.5
Mindfulness	16.1
Mindfulness	12.7
Mindfulness	14.9
Mindfulness	11.3

The exercise also asks you to add a `gender` variable using `sample()`.

The task (same as the slide exercise)

1. Add a `gender` variable to the data (use `sample()` to randomly assign "Male", "Female", "Non-binary")
2. Create a visualization showing depression scores by gender
3. Try at least two different geom types
4. Add appropriate labels

Your pen-and-paper version

Step 1: Think about the `mutate` + `sample` step. What would this code do?

```
therapy_data |>
  mutate(gender = sample(c("Male", "Female", "Non-binary"),
                         size = n(), replace = TRUE))
```

Why do we need `replace = TRUE`? _____

Why are we using `sample()` instead of real gender data? _____

Step 2: Plan two visualizations. For comparing a continuous variable (`depression_post`) across groups (`gender`), which geom types make sense? List at least 3 options and star the two you'd choose:

- `geom_____________`
- `geom_____________`
- `geom_____________`
- `geom_____________`

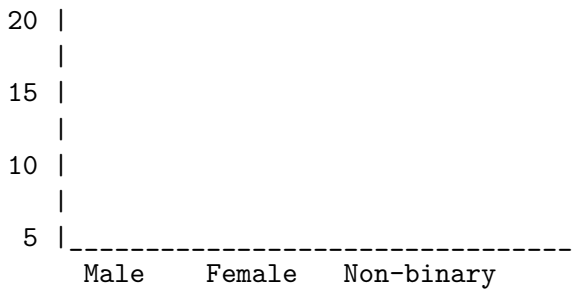
Step 3: Sketch two plots. Draw two different visualizations of depression scores by gender using the grid below.

Plot A (geom type: _____)

```
depression
25 |
   |
20 |
   |
15 |
   |
10 |
   |
 5 |_________________________________
     Male    Female   Non-binary
```

Plot B (geom type: _____)

```
depression
25 |
   |
```



Step 4: Write the code for one of your plots. Fill in the blanks:

```
ggplot(therapy_data_gender, aes(x = _______, y = _________________)) +
  geom_____________() +
  labs(
    title = "_________________________________",
    x = "____________",
    y = "____________"
  ) +
  theme_minimal()
```

Step 5: Think about it. Since we randomly assigned gender (it's not real data), would you expect to see a real difference between groups? Why or why not?

Your answer: _____

Check your work

Compare your sketches and code with your partner's screen. Do your geom choices and code match?