

Session 2: Your First Visualization — Pen-and-Paper Pair Exercise

PSY 410 | Data Science for Psychology

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: mpg

This dataset has fuel economy data for 234 cars. Here are 8 rows with the columns you'll need:

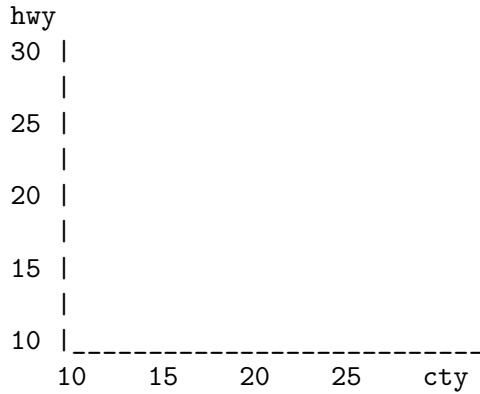
manufacturer	model	cty	hwy	fl
audi	a4	18	29	p
audi	a4	21	29	p
chevrolet	k1500 tahoe 4wd	14	17	d
chevrolet	c1500 suburban 2wd	11	15	e
chevrolet	c1500 suburban 2wd	14	20	r
jeep	grand cherokee 4wd	17	22	d
chevrolet	k1500 tahoe 4wd	11	14	e
chevrolet	c1500 suburban 2wd	14	20	r

Key: cty = city miles per gallon, hwy = highway miles per gallon, fl = fuel type (d = diesel, e = ethanol, p = premium, r = regular)

The task (same as the slide exercise)

1. Plot cty (x-axis) vs hwy (y-axis)
2. Color points by **fuel type** (fl)
3. Add a smooth trend line
4. Give it a title and axis labels

Step 1: Sketch the scatterplot. Use the grid below. Label the x-axis “City MPG” and the y-axis “Highway MPG.” Plot each of the 8 rows as a point. Use different symbols for each fuel type (e.g., circles for p, squares for r, triangles for d, stars for e) since you can’t use color on paper.



Step 2: Write the code. Even without a laptop, you can write the ggplot code that would produce this plot. Fill in the blanks:

```
ggplot(_____, aes(x = _____, y = _____, color = _____)) +
  geom_____() +
  geom_____() +
  labs(
    title = "_____",
    x = "_____",
    y = "_____",
    color = "_____ "
  )
```

Step 3: Predict. Look at the data. What pattern do you expect to see in the scatterplot? (Do cars with higher city MPG also tend to have higher highway MPG?)

Your prediction: