Quiz 1

You have 30 minutes to complete this 10 question quiz. The questions are weighted equally. You can consult any course materials or the internet. However, **you cannot use R and you must complete the quiz individually.**

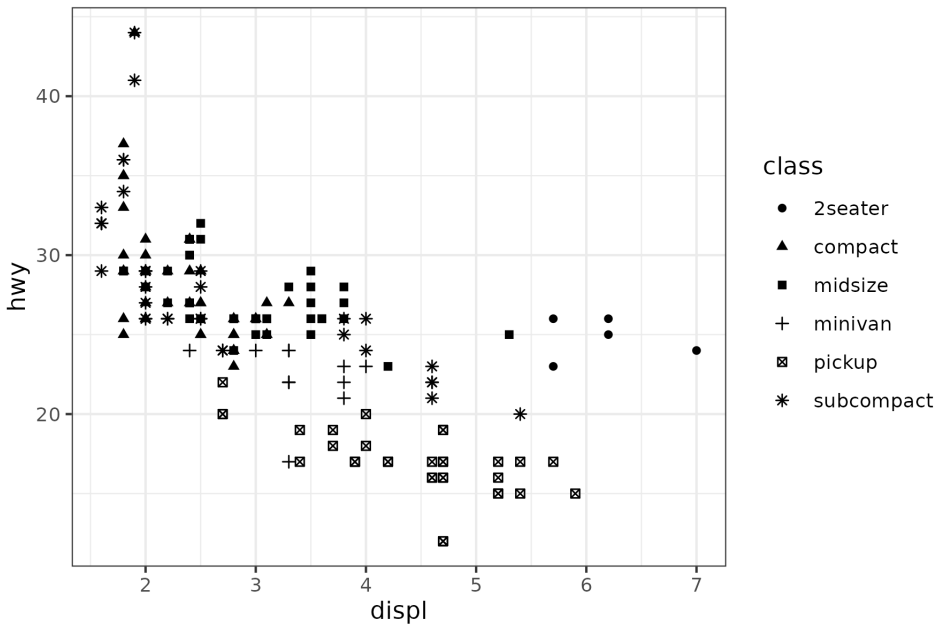
1 10 points

The next four questions are on the mpg data, several rows and columns of which are shown below:

A tibble: 234 × 5

displ <dbl>	year <int>	cyl <int>	hwy <int>	class <chr>
1.8	1999	4	29	compact
1.8	1999	4	29	compact
2.0	2008	4	31	compact
2.0	2008	4	30	compact
2.8	1999	6	26	compact
2.8	1999	6	26	compact
3.1	2008	6	27	compact
1.8	1999	4	26	compact
1.8	1999	4	25	compact
2.0	2008	4	28	compact

Consider the following ggplot:



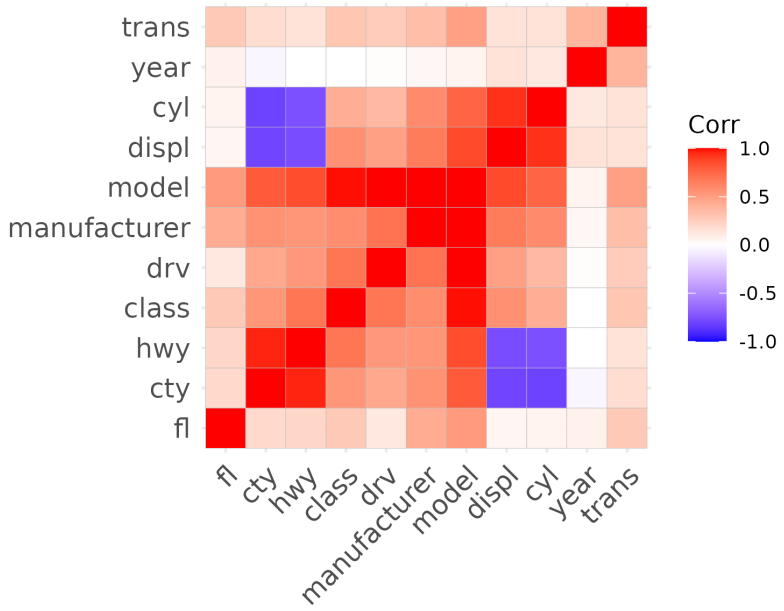
This plot was created using the following code (the arguments to aes are omitted):

```
ggplot(data = mpg) +
  geom_point(mapping = aes(...))
```

How many distinct aesthetics are mapped inside aes()? Enter your answer in numeric form.

2 10 points

Below is the correlation matrix for the mpg data. Which variable has the weakest relationship with hwy?



- fl
- cty
- class
- drv
- manufacturer
- model
- displ
- cyl
- year
- trans

3 10 points

Which of the following settings of an aesthetic should go inside of aes()? Select all that apply.

- [ ] color = class
- [ ] shape = "square"
- [ ] x = displ
- [ ] y = 3
- [ ] alpha = hwy

4 10 points

Which of the following code chunks outputs the number of compact cars with four cylinders? Select all that apply.

```
# Code chunk A
mpg %>%
  filter(class == "compact") %>%
  summarise(output = sum(cyl == 4))

# Code chunk B
mpg %>%
  summarise(output = sum(cyl == 4 & class == "compact"))

# Code chunk C
mpg %>%
  filter(cyl == 4) %>%
  summarise(output = all(class == "compact"))

# Code chunk D
mpg %>%
  filter(cyl == 4 & class == "compact") %>%
  summarise(output = n())
```

- [ ] A
- [ ] B
- [ ] C
- [ ] D

5 10 points

Consider the following two datasets:

```{r}
flights
```

A tibble: 8 × 8

origin <chr>	year <int>	month <int>	day <int>	hour <dbl>	flight <int>	carrier <chr>	distance <dbl>
EWR	2013	1	1	10	75	US	2133
LGA	2013	1	1	10	1847	DL	762
PHL	2013	1	1	10	2379	DL	1076
LGA	2013	1	1	10	1177	US	544
EWR	2013	1	1	10	503	B6	1065
PHL	2013	1	1	10	1625	UA	1605
EWR	2013	1	1	10	3795	MQ	719
LGA	2013	1	1	10	2319	DL	1020

```{r}
weather
```

A tibble: 3 × 8

origin <chr>	year <int>	month <int>	day <int>	hour <int>	temp <dbl>	precip <dbl>	pressure <dbl>
EWR	2013	1	1	10	41.00	0	1012.4
JFK	2013	1	1	10	41.00	0	1012.8
LGA	2013	1	1	10	39.92	0	1012.1

Suppose we join the datasets as follows:

```
left_join(flights, weather, by = c("origin", "year", "month", "day", "hour"))
```

How many NA values will the joined data contain? Enter your answer in numeric form.

6 10 points

Which dplyr verbs can return a dataset with a different number of rows than its input? Select all that apply.

- [ ] filter
- [ ] select
- [ ] arrange
- [ ] mutate
- [ ] summarise

7 10 points

Which dplyr verbs can return a dataset with a different number of columns than its input? Select all that apply.

- [ ] filter
- [ ] select
- [ ] arrange
- [ ] mutate
- [ ] summarise

8 10 points

The next three questions refer to the following tibble:

country <chr>	capital <chr>	metric <chr>	value <dbl>
USA	Washington, DC	GDP	20.0
China	Beijing	GDP	15.0
USA	Washington, DC	population	330.0
China	Beijing	population	1400.0
Canada	Ottawa	area	3.8
Canada	Ottawa	population	38.0

When tidied, it will contain _____ rows and _____ columns.

9 10 points

Which of the following code chunks would correctly tidy this data?

```
# Code chunk A
data %>%
  pivot_wider(names_from = "metric", values_from = "value")

# Code chunk B
data %>%
  pivot_wider(cols = -c("country", "capital"), names_from = "metric", values_from = "value")

# Code chunk C
data %>%
  pivot_longer(names_to = "metric", values_to = "value")

# Code chunk D
data %>%
  pivot_longer(cols = -c("country", "capital"), names_to = "metric", values_to = "value")
```

- A
- B
- C
- D
- None of the above

10 10 points

How many missing values would the tidied version of the data contain? Enter your answer in numeric form.