# *Starting with Graphics and Wrangling*

## *Data Computing*

In today's activity, you are going to deconstruct some graphics and carry out some data wrangling operations.[1]

### *Deconstructing graphics*

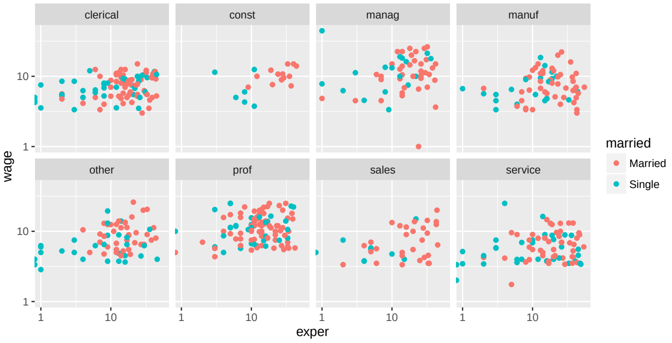

Figure 1: A representation of some of the variables from the CPS85 data table in the mosaicData package.

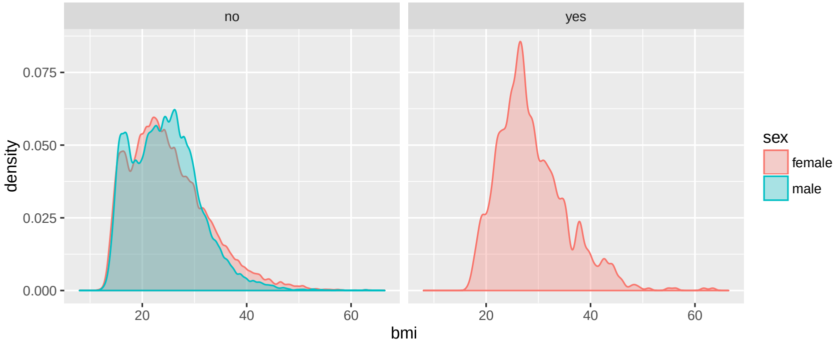

Figure 2: Variables from the NCHS data table in the DataComputing package. The 'yes' and 'no' refers to whether the person is pregnant.

Considering each of the above graphics in turn, figure out:

- What mode of graphic is it? (e.g. density plot, scatter plot, bar plot, . . . )
- What variables from the respective data tables are involved?
- What role each of those variables plays in the graphic?
- In Figure 2, why is there no data variable being used for the $y$-axis?

Here is the basic structure of the commands for making the graphics. You can try various combinations of the variables appearing in the graphics and see which graphic you think is the most informative.

[1] Remember to load the `DataComputing` package: `library(DataComputing)`

```
ggplot(data = CPS85, aes(x = ????, y = ????, color = ????)) + geom_point() + facet_wrap( ~ ????)
ggplot(data = NCHS, aes(x = ????)) + geom_density(aes(color = ????)) + facet_wrap(~ ???)
```

Put the R statement that generates each graph into your report so that the graphs appear when you compile your `.Rmd` file.

## *Wrangling*

### Diamonds

Refer to the `diamonds` data table in the `ggplot2` package. Take a look at the codebook (using `help()`) so that you'll understand the meaning of the tasks. (Motivated by Garrett Grolemund.)

Each of the following tasks can be accomplished by a statement of the form

```
diamonds %>%
  verb1( args1 ) %>%
  verb2( args2 ) %>%
  arrange(desc( args3 )) %>%
  head( 1 )
```

For each task, give appropriate R functions or arguments to substitute in place of `verb1`, `verb2`, `args1`, `args2`, and `args3`.

1. Which color diamonds seem to be largest on average (in terms of carats)?
2. Which clarity of diamonds has the largest average "table" per carat?

---

### Voting

Using the `Minneapolis2013` data table, answer these questions:

`summarise()`, `group_by()`, and `tally()` will be useful in answering the questions.

1. How many cases are there?
2. Who were the top 5 candidates in the `Second` vote selections.
3. How many ballots are marked "undervote" in
   - `First` choice selections?
   - `Second` choice selections?
   - `Third` choice selections?
4. What are the top 3 `Second` vote selections among people who voted for Betsy Hodges as their first choice?
5. Which `Precinct` had the highest fraction of `First` vote selections marked as "undervote"?

To calculate the fraction, use `mean(First == "undervote")` in the argument to `summarise()`.