# R Workshop Day 3

## Data manipulation with dplyr and the tidyverse

- The tidyverse family of packages[1]
- data manipulations using dplyr[2]

# Practice

1. Enhance the graph we generated for femalesByState by adding a reference line at the nationwide female percentage. To compute that, you will need to do some more work using summarize on the femalesByState dataset, to compute the total female population as well as the total population, then divide the two.
2. Add a text next to the reference line from the previous part, indicating the value.
3. Repeat a similar analysis by using the high school graduation percentage instead of the female percentage.
4. Compute both black and female population percentages per state, then draw an scatterplot comparing these two variables.
   - Also compute the nationwide black percentage and female percentage and add appropriate horizontal and vertical reference lines at those locations (preferably red).
   - Add state labels for those states that are below the nationwide average in both.
   - Arrange the states dataset in order of female percentage, then use gf\_line to join all the points together (maybe with some alpha).
   - Also include a computation of high school graduation percentages, and color the points (only the points) using that.

---

[1]../morsels/tidyverse.html
[2]../morsels/dplyr.html