

Practice Problems on Scatterplots and Correlation

1. Load the data set `health.csv` available [here](#).
2. Use the `glimpse()` function in the `dplyr` package to get a quick overview of the data. If you need more information, see the codebook [here](#).
3. Use the `plot()` function to create a scatterplot of `ideology` and `health_score`. Higher values of the variable `ideology` indicate a more conservative state. The variable `health_score` is an overall summary of the healthiness in the state and that higher values indicate a more healthy state. Make sure to specify the `xlab` and `ylab` arguments so the plot looks really nice. If you want, you can also give the plot a title using the `main` argument.
4. Use the `cor()` function to calculate the correlation between `ideology` and `health_score`. Is it consistent with the scatterplot?
5. Repeat questions 3 and 4 using `cancer_incidence`. Do you run into any hiccups? How might you solve this problem? Hint: Open the help file for `cor()` and carefully look at the arguments. Does one seem helpful? Are the results consistent with the overall health score?
6. Repeat questions 3 and 4 with another measure of state health, such as the life expectancy or infant mortality rate. Are the results still consistent?
7. What does this tell us about health politics? Does it suggest optimism or pessimism about the ability of state government intervention to help solve health issues in the US?
8. How do you interpret a correlation coefficient? What does -1 mean? 1? 0? What can you say about a correlation of 0.5 compared to a correlation of 0.3? The figure below provides some helpful examples.

