Correlation

Prepared by Gerald Schafer, June 2016

While it is true that correlation does not equal causation, investigating correlations may reveal associations between variables that warrant further study.

[CDC Wonder](http://wonder.cdc.gov) is a great source of freely available health data. The goal of this project is to explore CDC Wonder and to brainstorm and investigate some interesting relationships that can be informally tested through plotting of correlations. Note that we are not formally testing hypotheses.

Example: Sun exposure is a [known risk factor](http://www.cancer.org/cancer/skincancer-melanoma/detailedguide/melanoma-skin-cancer-risk-factors) for melanoma skin cancer. Although there are other potential risk and protective factors for melanoma, it is plausible that there is a positive relationship on a state level between the level of solar radiation and the incidence of melanoma skin cancer in that state. With CDC Wonder, we can do a quick investigation of this this relationship. Here’s what it looks like.

Figure: Daily Sunlight and Melanoma Incidence by State

**Question**:

* How would you describe the relationship between daily sunlight and the incidence of melanoma?
* Does this graph surprise you or not? Why?
* Other than solar radiation, what other factors might influence the incidence of melanoma in a given state?
* The trendline and equation are included on the graph. What do they reveal?

Explore CDC Wonder and identify possible associations between two variables that you find interesting and for which there’s a plausible relationship. Explain the anticipated relationship and create a graph similar to the figure above. Evaluate the strength of the relationship between your factors. Discuss your findings, including unexpected results, and give a possible explanation for these findings.