

Cloud Cover

Anonymous

December 2, 2014

Data

The `atmos` data set resides in the `nasaweather` package of the *R* programming language. It contains a collection of atmospheric variables measured between 1995 and 2000 on a grid of 576 coordinates in the western hemisphere. The data set comes from the 2006 ASA Data Expo.

Some of the variables in the `atmos` data set are:

- **cloudlow** - The mean percent of the sky covered by clouds at low altitudes.
- **cloudmid** - The mean percent of the sky covered by clouds at mid-range altitudes.
- **cloudhigh** - The mean percent of the sky covered by clouds at high altitudes.

You can convert the temperature unit from Kelvin to Celsius with the formula

$$celsius = kelvin - 273.15$$

And you can convert the result to Fahrenheit with the formula

$$fahrenheit = celsius \times \frac{9}{5} + 32$$

Cleaning

For the remainder of the report, we will look only at data from the year 2000. We aggregate our data by location, using the *R* code below.

```
means <- atmos %>%
  filter(year == year) %>%
  group_by(long, lat) %>%
  summarize(temp = mean(temp, na.rm = TRUE),
            pressure = mean(pressure, na.rm = TRUE),
            ozone = mean(ozone, na.rm = TRUE),
            cloudlow = mean(cloudlow, na.rm = TRUE),
            cloudmid = mean(cloudmid, na.rm = TRUE),
            cloudhigh = mean(cloudhigh, na.rm = TRUE)) %>%
  ungroup()

clouds <- means %>%
  select(-(temp:ozone)) %>%
  gather("altitude", "coverage", 3:5)
```