

Ukázka 2

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:

- **temp** - The mean monthly air temperature near the surface of the Earth (measured in kelvins (*K*))
- **pressure** - The mean monthly air pressure at the surface of the Earth (measured in millibars (*mb*))
- **ozone** - The mean monthly abundance of atmospheric ozone (measured in Dobson units (*DU*))

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

$$celsius = kelvins - 273.15$$

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 1995. We aggregate our data by location, using the *R* code below.

```
library(nasaweather)
library(dplyr)
library(ggvis)

year <- 1995

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()
```

We suspect that group level effects are caused by environmental conditions that vary by locale. To test this idea, we sort each data point into one of four geographic regions:

```
means$locale <- "north america"
means$locale[means$lat < 10] <- "south pacific"
means$locale[means$long > -80 & means$lat < 10] <- "south america"
means$locale[means$long > -80 & means$lat > 10] <- "north atlantic"
```

```
lm(ozone ~ temp + locale + temp:locale, data = means)
```

```
##
```

```
## Call:
```

```
## lm(formula = ozone ~ temp + locale + temp:locale, data = means)
```

```
##
```

```
## Coefficients:
```

```
##           (Intercept)                temp  
##           1336.508                -3.559  
##    localenorth atlantic    localesouth america  
##           548.248                -1061.452  
##    localesouth pacific temp:localenorth atlantic  
##           -549.906                -1.827  
## temp:localesouth america temp:localesouth pacific  
##           3.496                    1.785
```