# Temperature Plots

*Tom Gocken*

*Tuesday, January 26, 2016*

## Load previously created tidy dataset, envdat

```
setwd("C:/Projects/springboard-wrangling")
envdat <- read.delim("../data/envdat.txt")
```

## Create by month and by year datasets

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
envdat$yr_mo <- envdat$year + envdat$month / 12

envdat_by_month <- envdat %>%
  group_by(yr_mo) %>%
  summarize(max_air_temp_mean = mean(max_air_temp),
            max_air_temp_median = median(max_air_temp),
            n = n()) %>%
  arrange(yr_mo)

envdat_by_year <- envdat %>%
  group_by(year) %>%
  summarize(max_air_temp_mean = mean(max_air_temp),
            max_air_temp_median = median(max_air_temp),
            n = n()) %>%
  arrange(year)
```
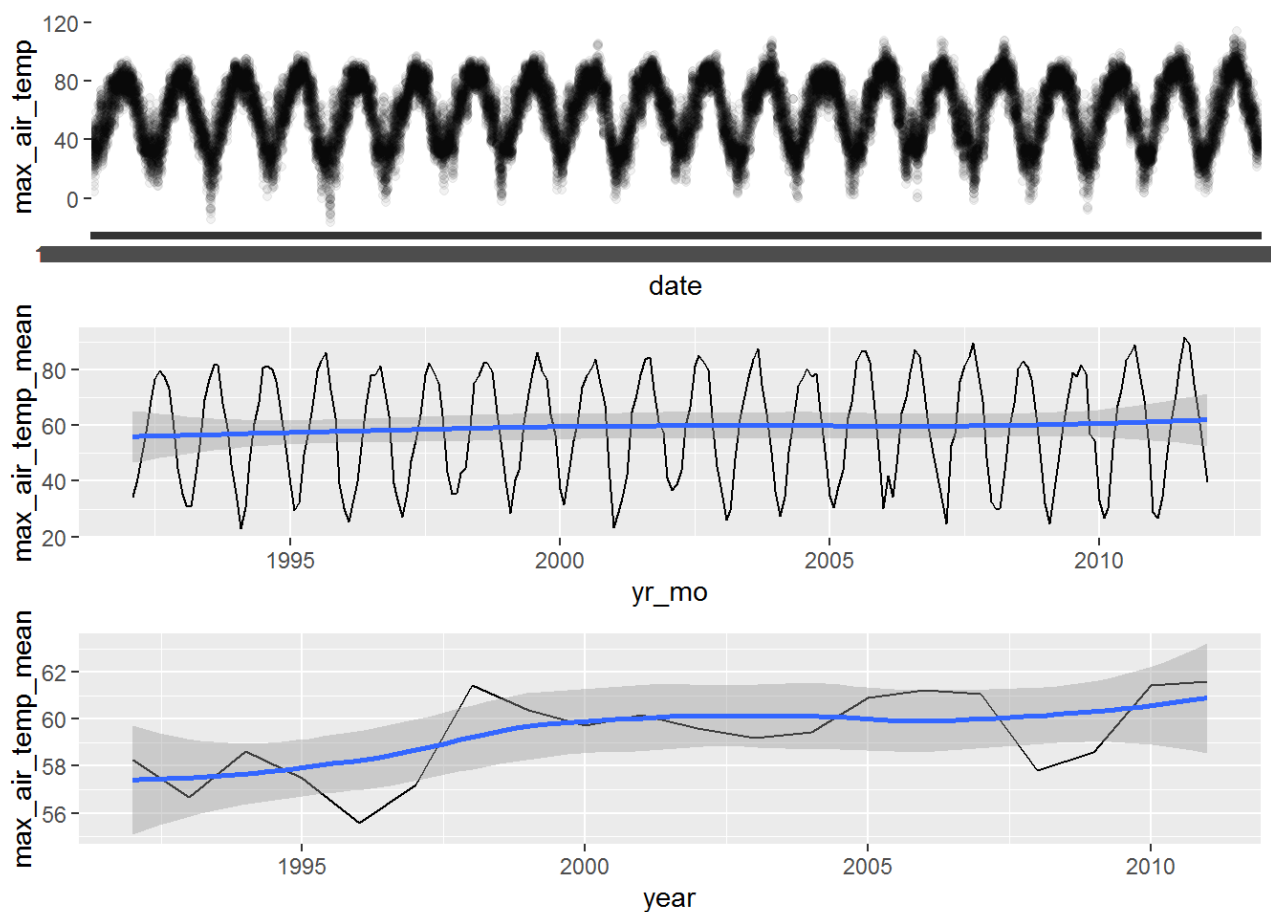
# Plot raw max temp air data, by month data, and by year data

```
library(ggplot2)
library(gridExtra)

p1 <- ggplot(aes(x = date, y = max_air_temp), data = envdat) + geom_point(alph
a = 1/20)
p2 <- ggplot(aes(x = yr_mo, y = max_air_temp_mean), data = envdat_by_month) +
  geom_line() + geom_smooth()
p3 <- ggplot(aes(x = year, y = max_air_temp_mean), data = envdat_by_year) +
  geom_line() + geom_smooth()

grid.arrange(p1, p2, p3, ncol = 1)
```



# Correlations of max air temp with month and year

```
cor.test(envdat_by_month$yr_mo, envdat_by_month$max_air_temp_mean, method = 'pe
arson')
```

```
##
##  Pearson's product-moment correlation
##
## data:  envdat_by_month$yr_mo and envdat_by_month$max_air_temp_mean
## t = 1.0033, df = 238, p-value = 0.3167
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.06224436  0.18996586
## sample estimates:
##        cor
## 0.06489707
```

```
cor.test(envdat_by_year$year, envdat_by_year$max_air_temp_mean, method = 'pearson')
```

```
##
##  Pearson's product-moment correlation
##
## data:  envdat_by_year$year and envdat_by_year$max_air_temp_mean
## t = 3.1131, df = 18, p-value = 0.006005
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2019319 0.8195574
## sample estimates:
##       cor
## 0.5915888
```