

# Week 1 - Notes

## Data Load

```
avgpm25_gitraw <- "https://raw.githubusercontent.com"
avgpm25_user <- "/jtleek/modules/master/04_ExploratoryAnalysis/exploratoryGraphs/data"
avgpm25_filename <- "/avgpm25.csv"
avgpm25 <- paste(avgpm25_gitraw, avgpm25_user, avgpm25_filename, sep = "")

download.file(avgpm25, destfile = "./data/avgpm25.csv")

classes <- c("numeric", "character", "factor", "numeric", "numeric")

pollution <- read.csv("./data/avgpm25.csv", colClasses = classes)
head(pollution)
```

```
##      pm25  fips region longitude latitude
## 1  9.771185 01003  east -87.74826 30.59278
## 2  9.993817 01027  east -85.84286 33.26581
## 3 10.688618 01033  east -87.72596 34.73148
## 4 11.337424 01049  east -85.79892 34.45913
## 5 12.119764 01055  east -86.03212 34.01860
## 6 10.827805 01069  east -85.35039 31.18973
```

```
summary(pollution$pm25)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3.383   8.549  10.047   9.836  11.356  18.441
```

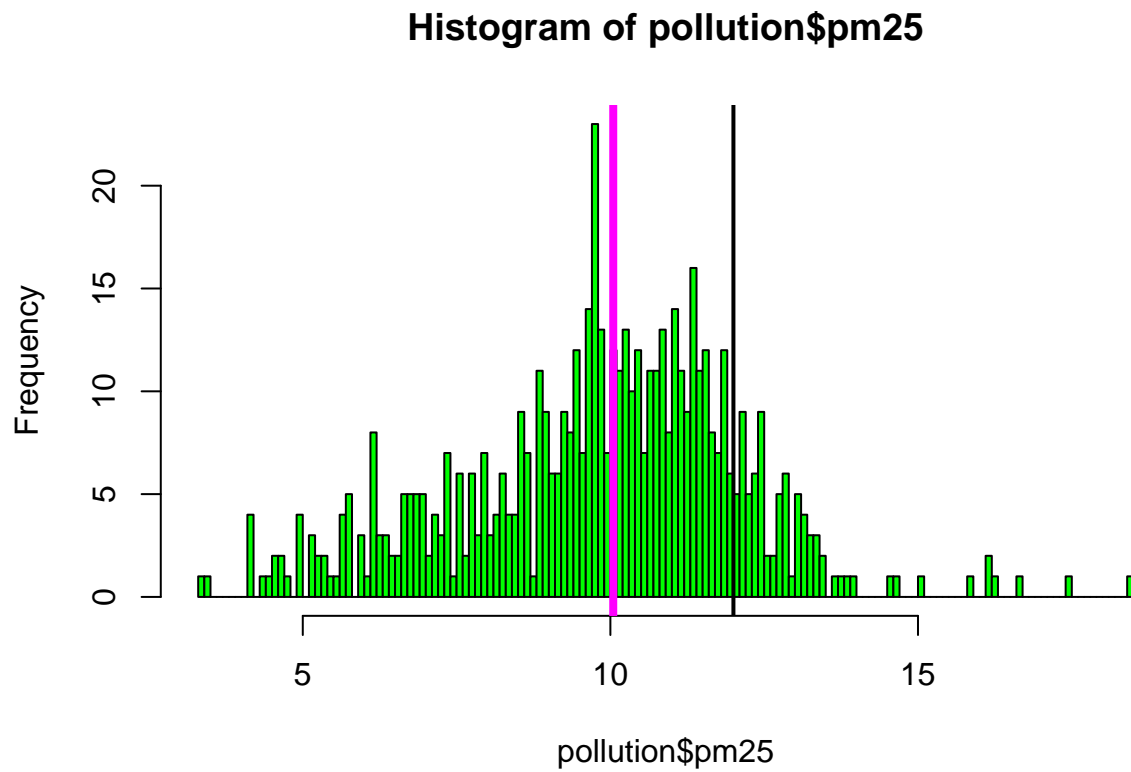
## Histograms

Simple histogram, with:

Vertical line at national limit of 12.

Vertical line at median.

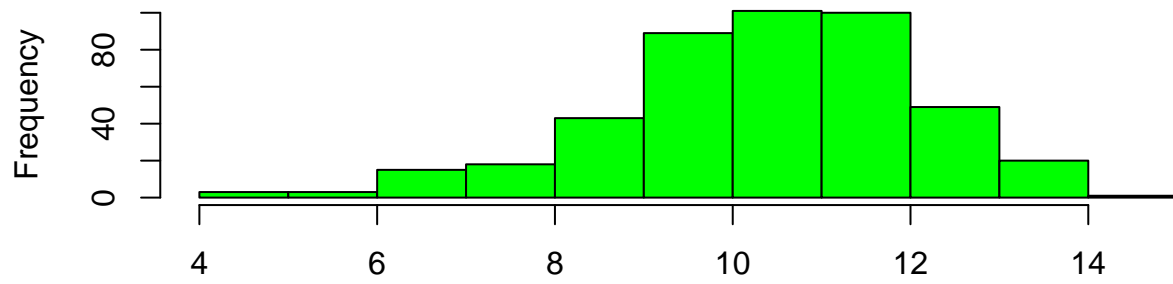
```
hist(pollution$pm25, col = "green", breaks=200)
abline(v = 12, lwd = 2)
abline(v = median(pollution$pm25), col = "magenta", lwd = 4)
```



#### Multiple

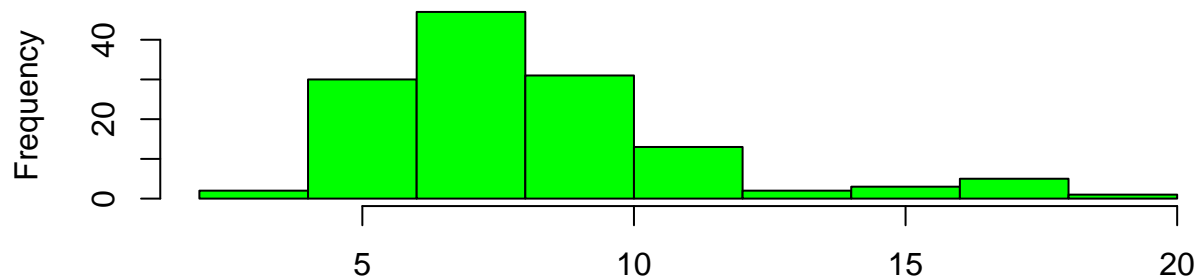
```
par(mfrow = c(2, 1), mar = c(4, 4, 2, 1))  
hist(subset(pollution, region == "east")$pm25, col = "green")  
hist(subset(pollution, region == "west")$pm25, col = "green")
```

**Histogram of subset(pollution, region == "east")\$pm25**



subset(pollution, region == "east")\$pm25

**Histogram of subset(pollution, region == "west")\$pm25**

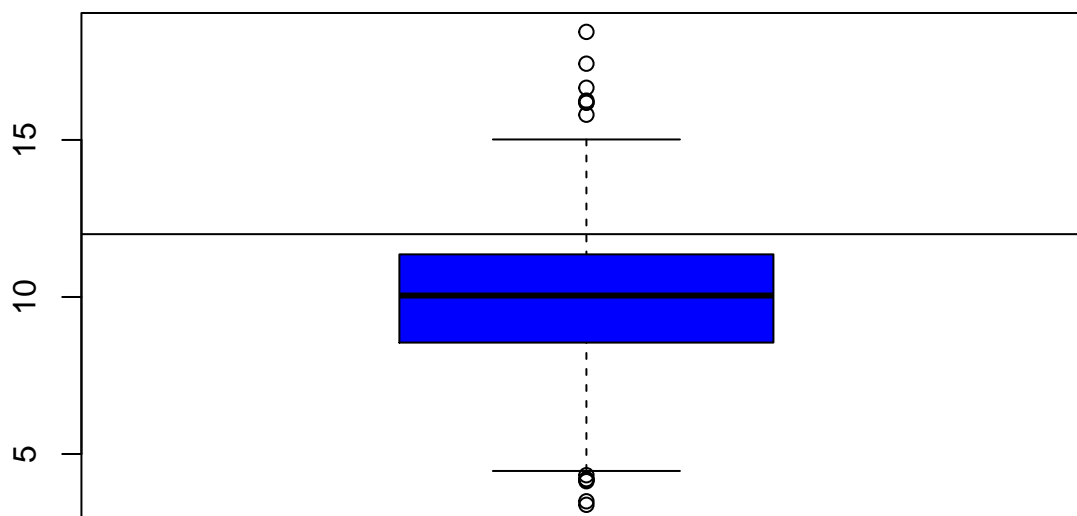


subset(pollution, region == "west")\$pm25

## Bloplots

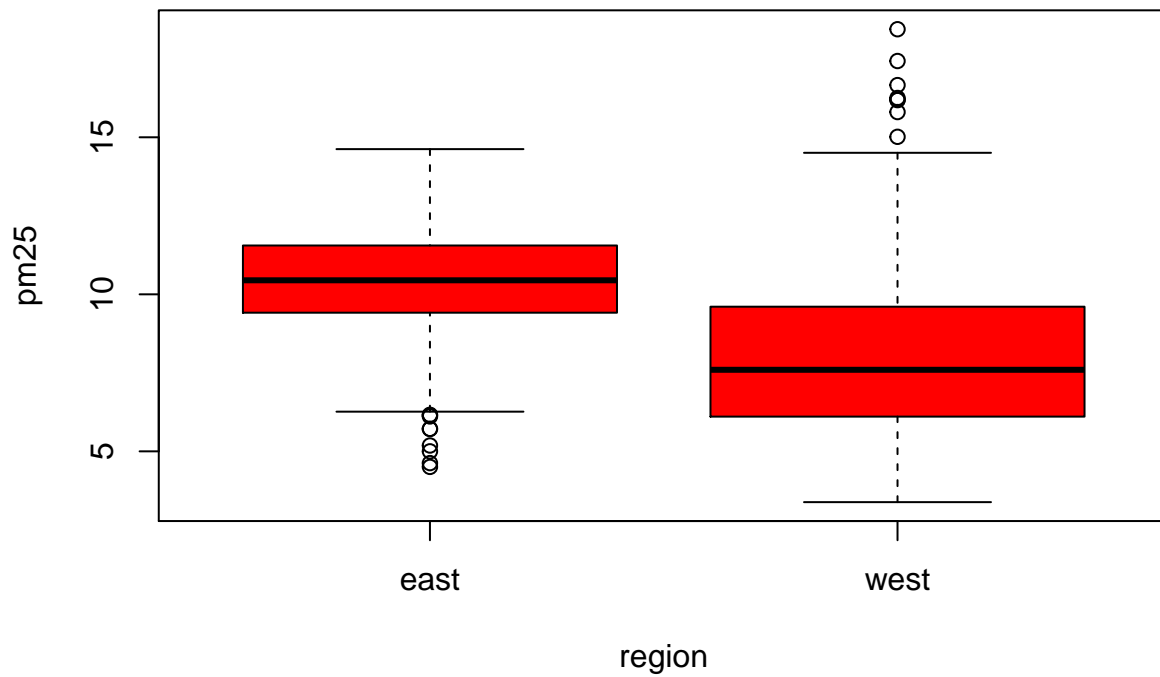
### Simple

```
boxplot(pollution$pm25, col = "blue")  
abline(h = 12)
```



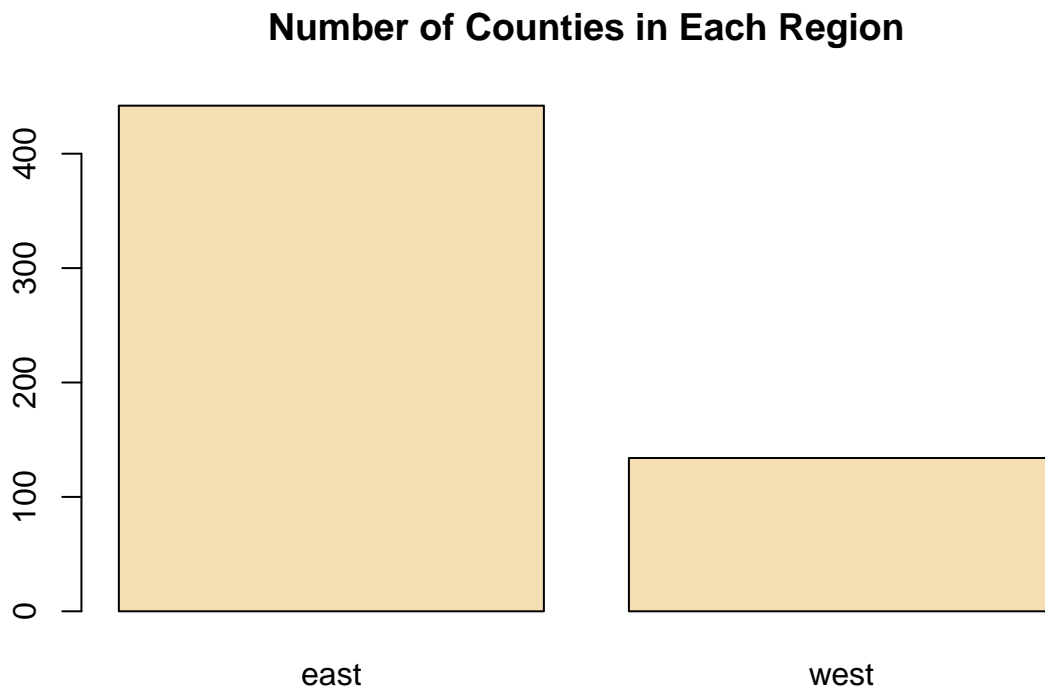
By region

```
boxplot(pm25 ~ region, data = pollution, col = "red")
```



Barplot

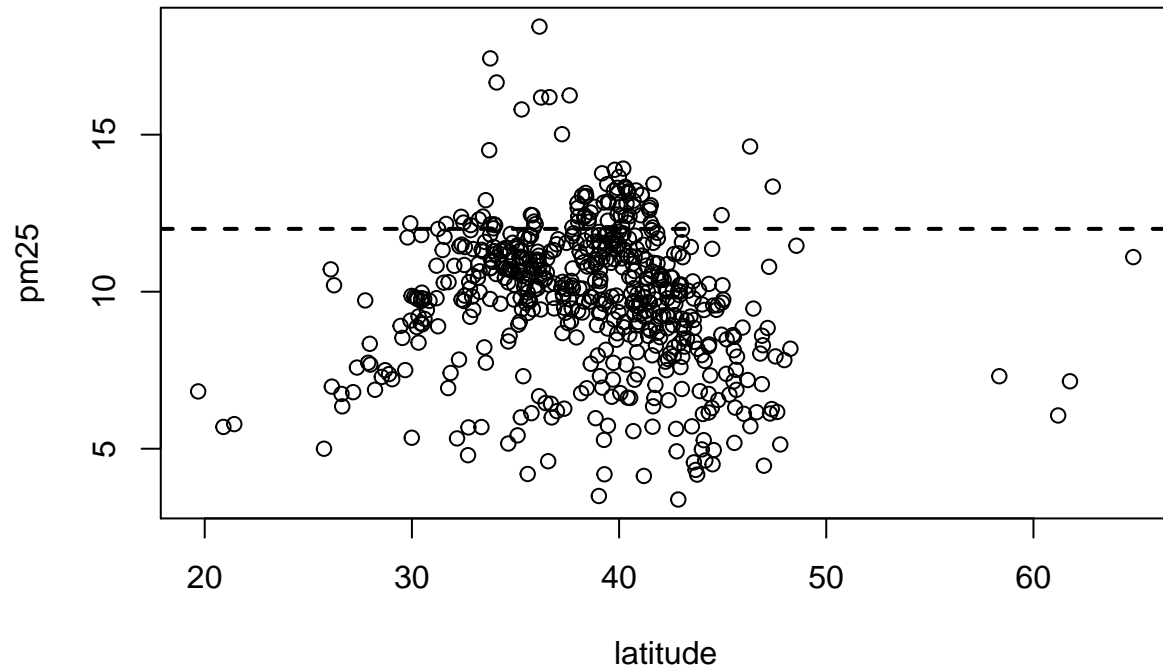
```
barplot(table(pollution$region), col = "wheat", main = "Number of Counties in Each Region")
```



# Scatterplots

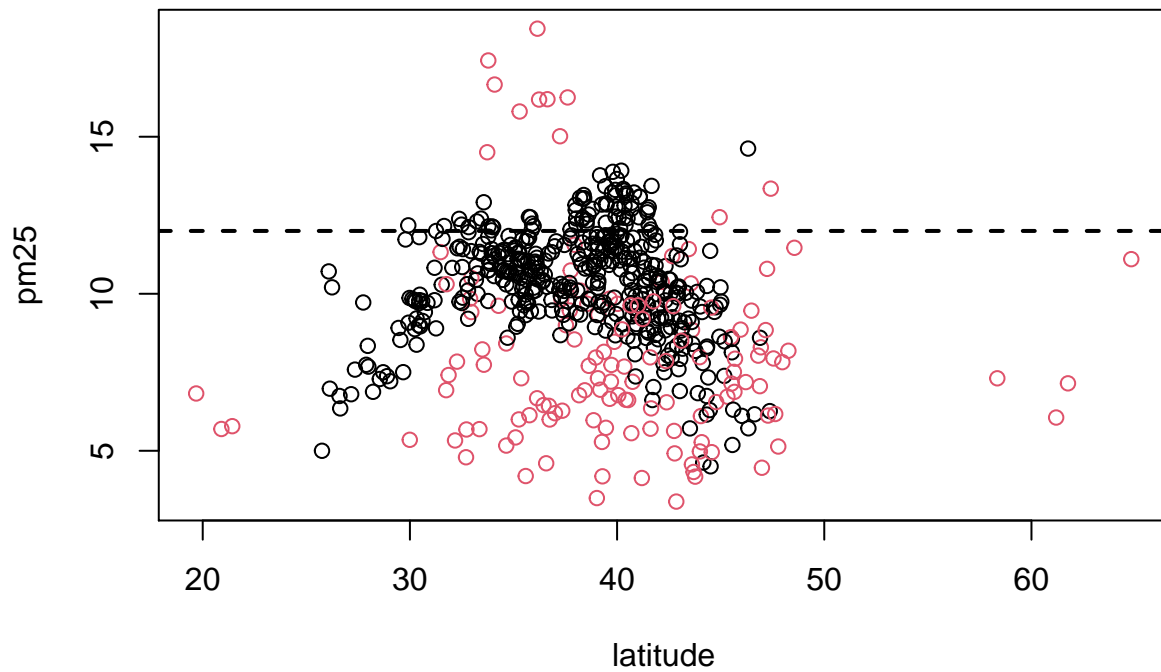
## Simple

```
with(pollution, plot(latitude, pm25))  
abline(h=12,lwd=2,lty=2)
```



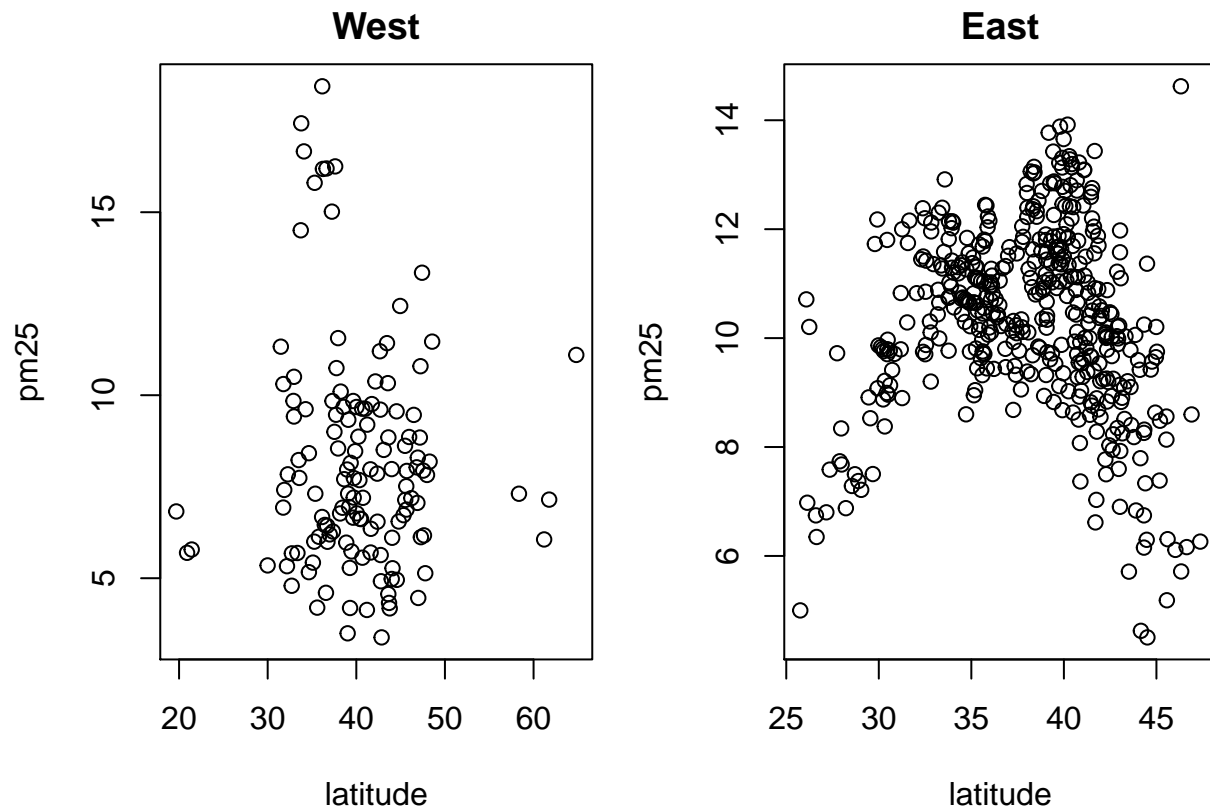
## With color

```
with(pollution, plot(latitude, pm25, col = region))  
abline(h=12,lwd=2,lty=2)
```



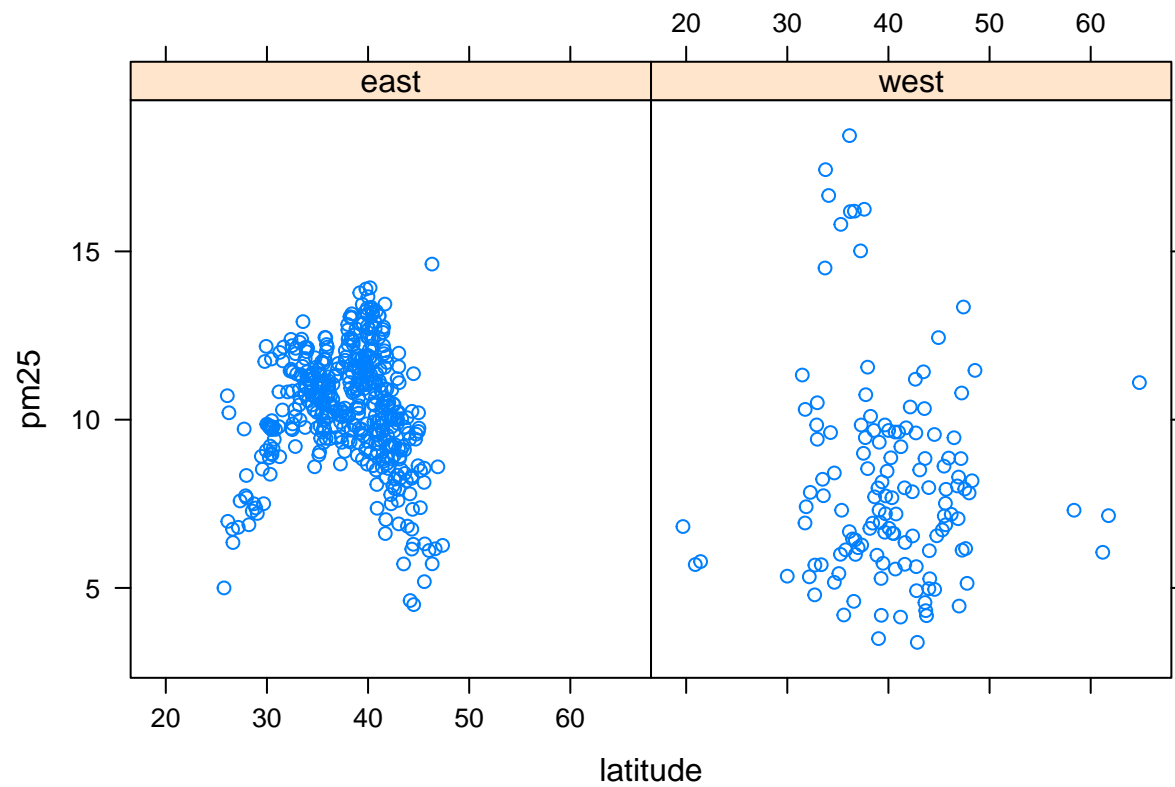
### Multiple

```
par(mfrow = c(1, 2), mar = c(5, 4, 2, 1))
with(subset(pollution, region == "west"), plot(latitude, pm25, main = "West"))
with(subset(pollution, region == "east"), plot(latitude, pm25, main = "East"))
```



## Lattice

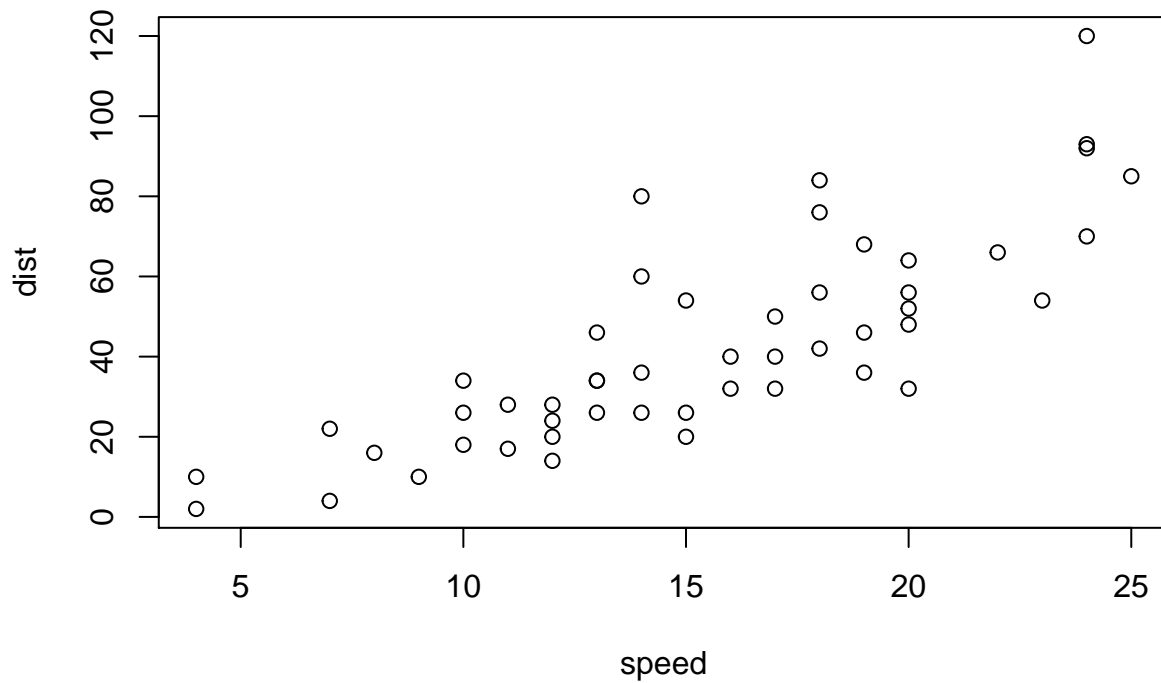
```
library(lattice)
xyplot(pm25 ~ latitude | region, data = pollution)
```



## Building Plots

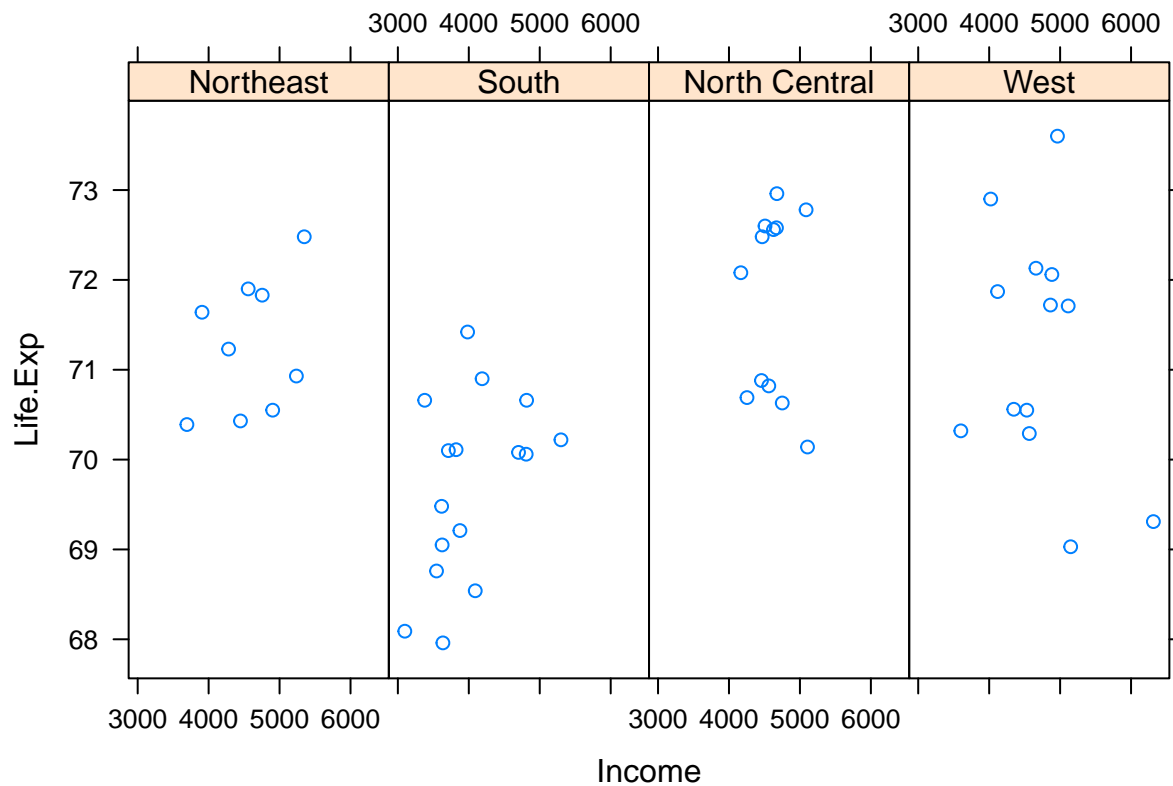
### Base Plotting System

```
library(datasets)
data(cars)
with(cars, plot(speed, dist))
```



## Lattice System

```
library(lattice)
state <- data.frame(state.x77, region = state.region)
xyplot(Life.Exp ~ Income | region, data = state, layout = c(4, 1))
```





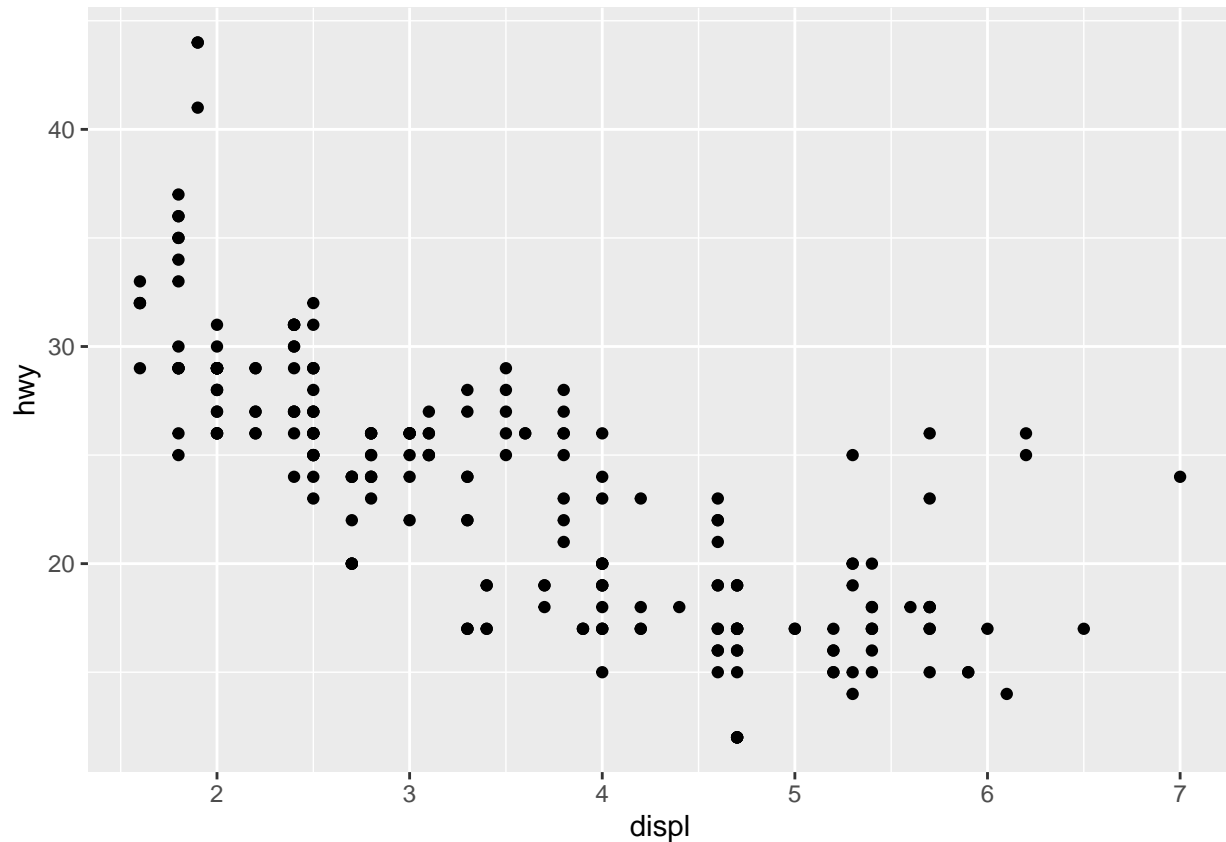
## ggplot2

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
data(mpg)
```

```
qplot(displ, hwy, data = mpg)
```



## Base Plots

### Base plot with annotation

#### With regression line

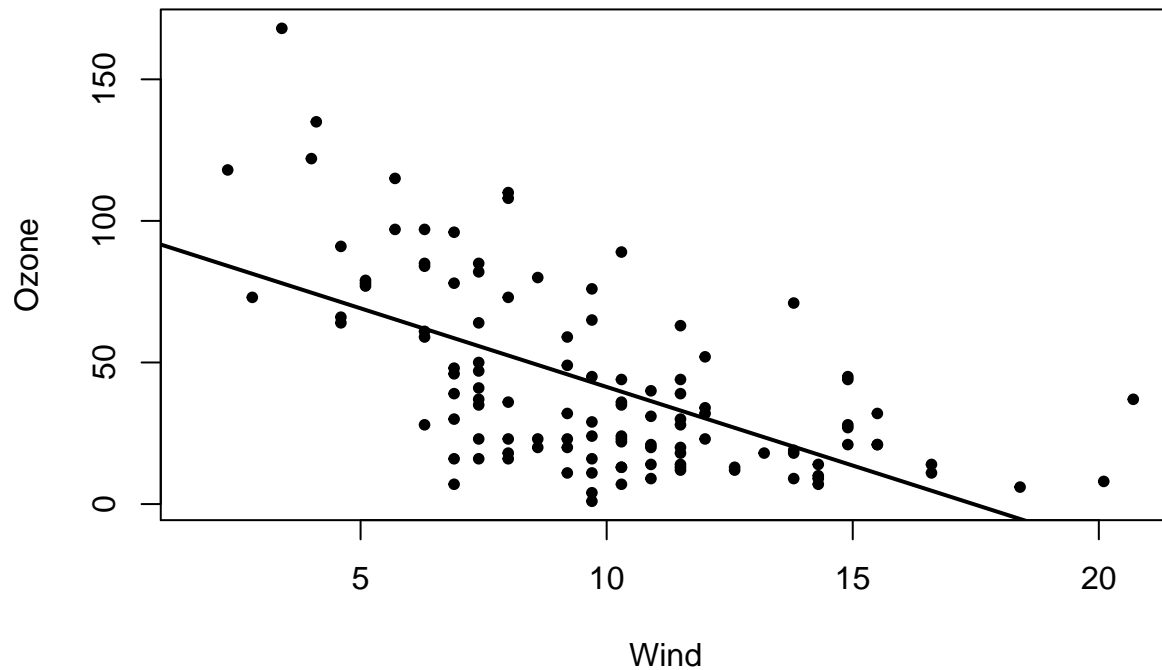
```
library(datasets)
```

```
with(airquality, plot(Wind, Ozone, main = "Ozone and Wind in New York City", pch = 20))
```

```
model <- lm(Ozone ~ Wind, airquality)
```

```
abline(model, lwd = 2)
```

## Ozone and Wind in New York City



Multiple

## Graphics Devices

where to put the plots: window, PDF, JPEG, etc

### PDF

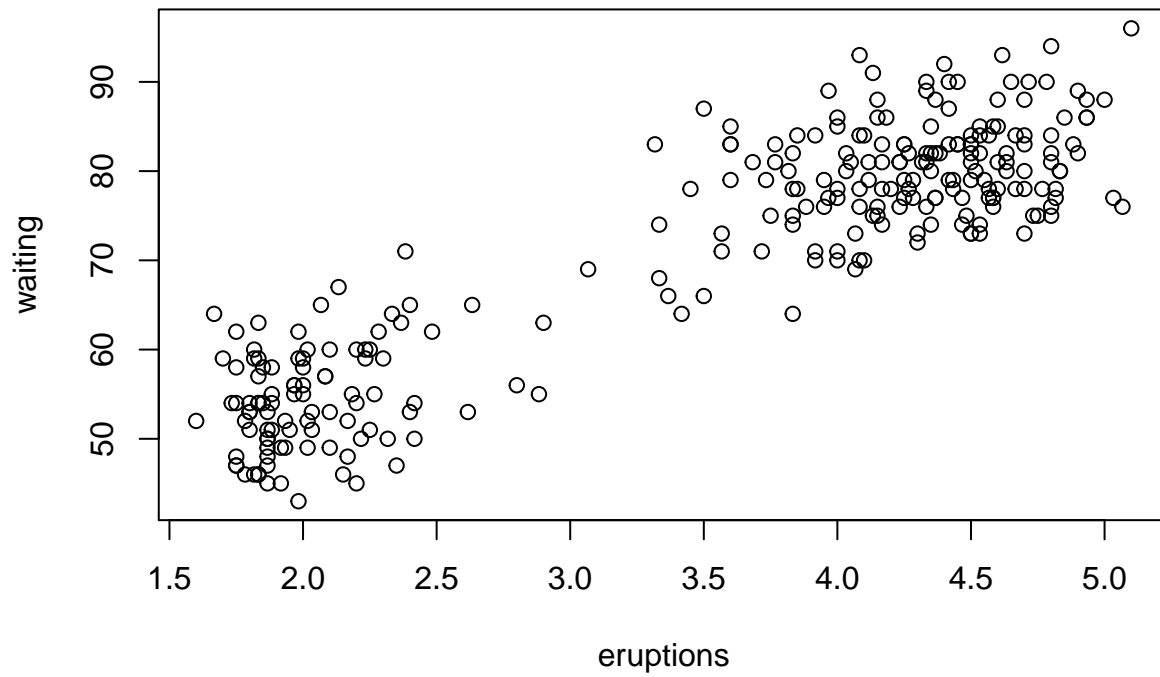
```
pdf(file = "myplot.pdf")
with(faithful, plot(eruptions, waiting))
title(main = "Old Faithful Geyser data")
dev.off()
```

```
## pdf
## 2
```

### PNG

```
library(datasets)
with(faithful, plot(eruptions, waiting))
title(main = "Old Faithful Geyser data")
```

## Old Faithful Geyser data



```
dev.copy(png, file = "geyserplot.png")
```

```
## quartz_off_screen  
## 3
```

```
dev.off()
```

```
## pdf  
## 2
```