

Advanced Visualization

Topics

- ggplot2
- Maps

ggplot2

ggplot2 developed by Hadley Wickham, based on the “grammar of graphics”

Particularly well suited for multi-dimensional, multivariate analysis.

Requires 3 things:

1. Data
2. Aesthetics
3. Layers

Data

```
library(raw)
```

```
## Warning: package 'raw' was built under R version 3.3.2
```

```
data(RegionExperience)  
dim(RegionExperience)
```

```
## [1] 40  4
```

```
head(RegionExperience)
```

```
##           Region PolicyYear NumPolicies NumClaims  
## 1 North Central      2001         579      4285  
## 2 North Central      2002         585      4459  
## 3 North Central      2003         588      4312  
## 4 North Central      2004         593      4377  
## 5 North Central      2005         598      4419  
## 6 North Central      2006         602      4518
```

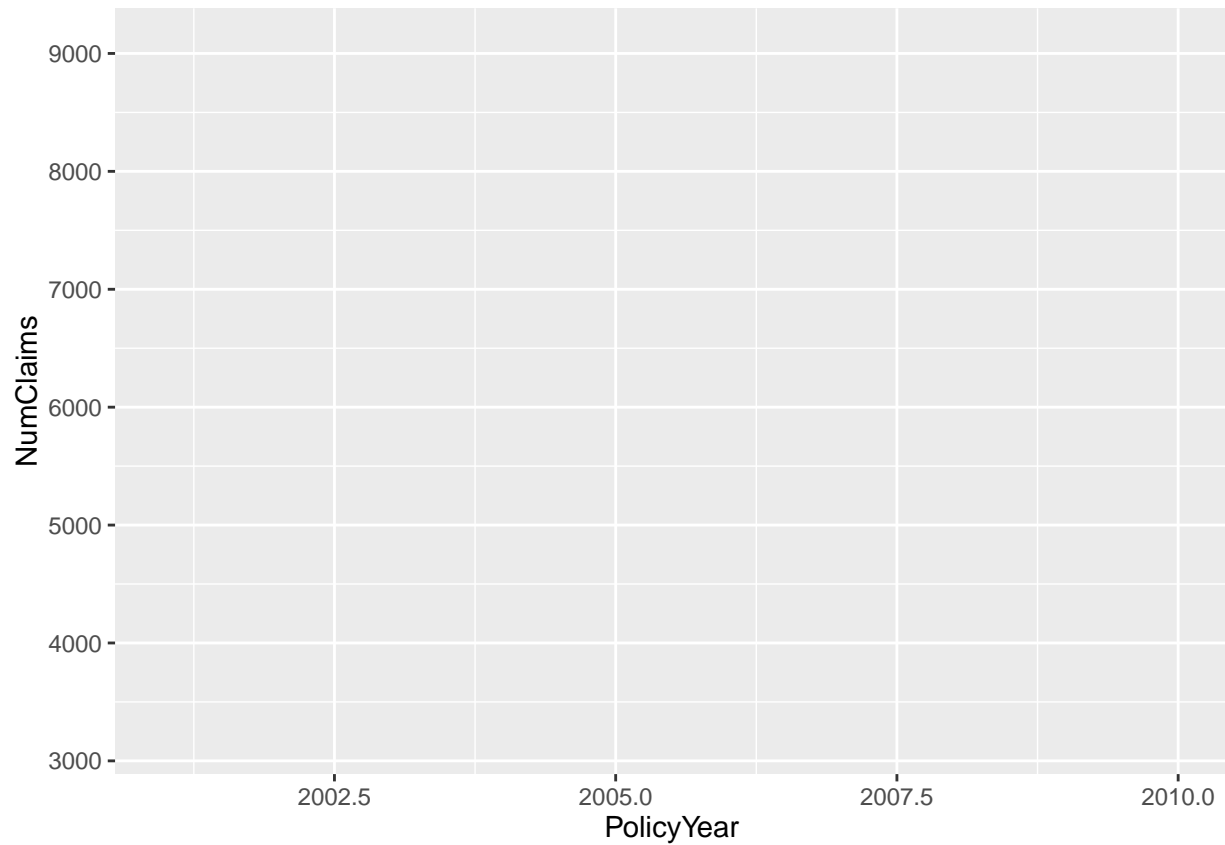
```
library(ggplot2)  
basePlot <- ggplot(RegionExperience)  
class(basePlot)
```

```
## [1] "gg"      "ggplot"
```

Aesthetics

Aesthetics are anything visible on the plot. Added with the `aes` function. Data will be mapped to visual elements.

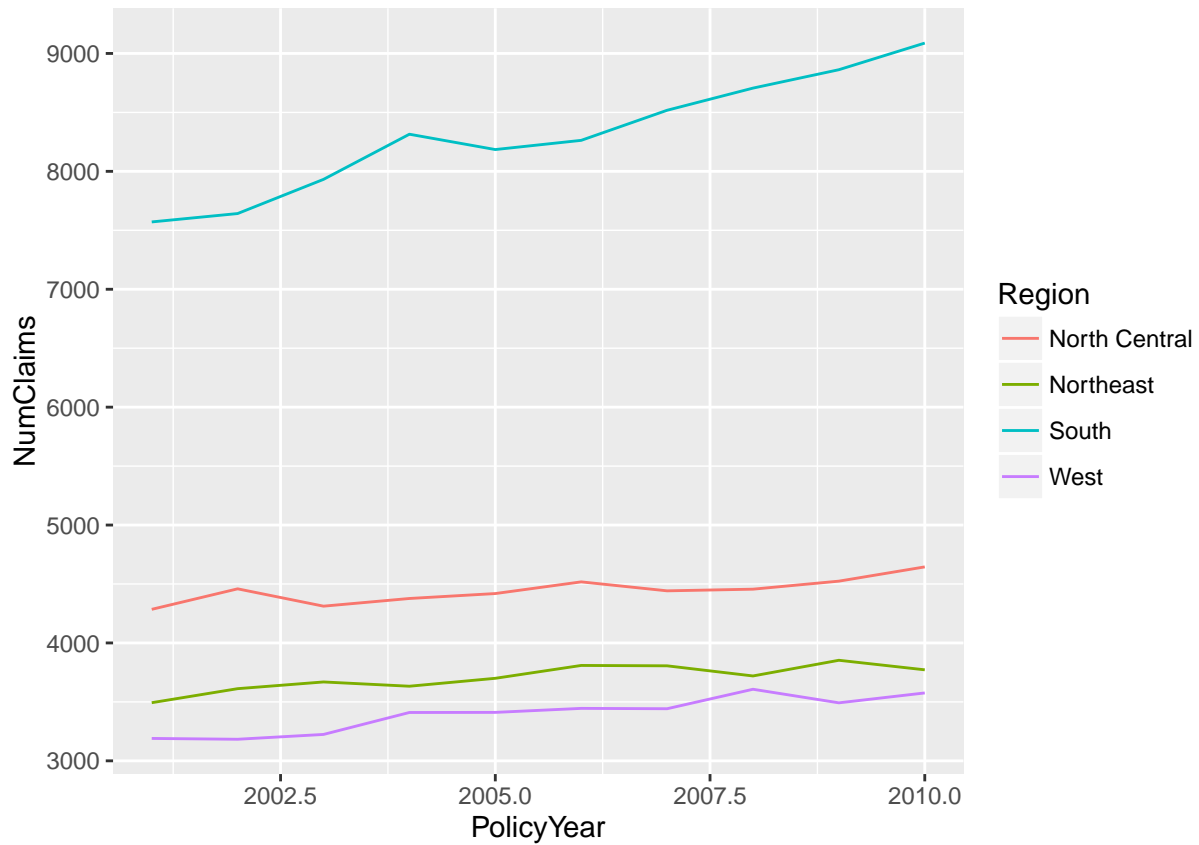
```
p <- basePlot + aes(x = PolicyYear, y = NumClaims, group=Region, color=Region, fill=Region)  
p
```



Adding layers: lines

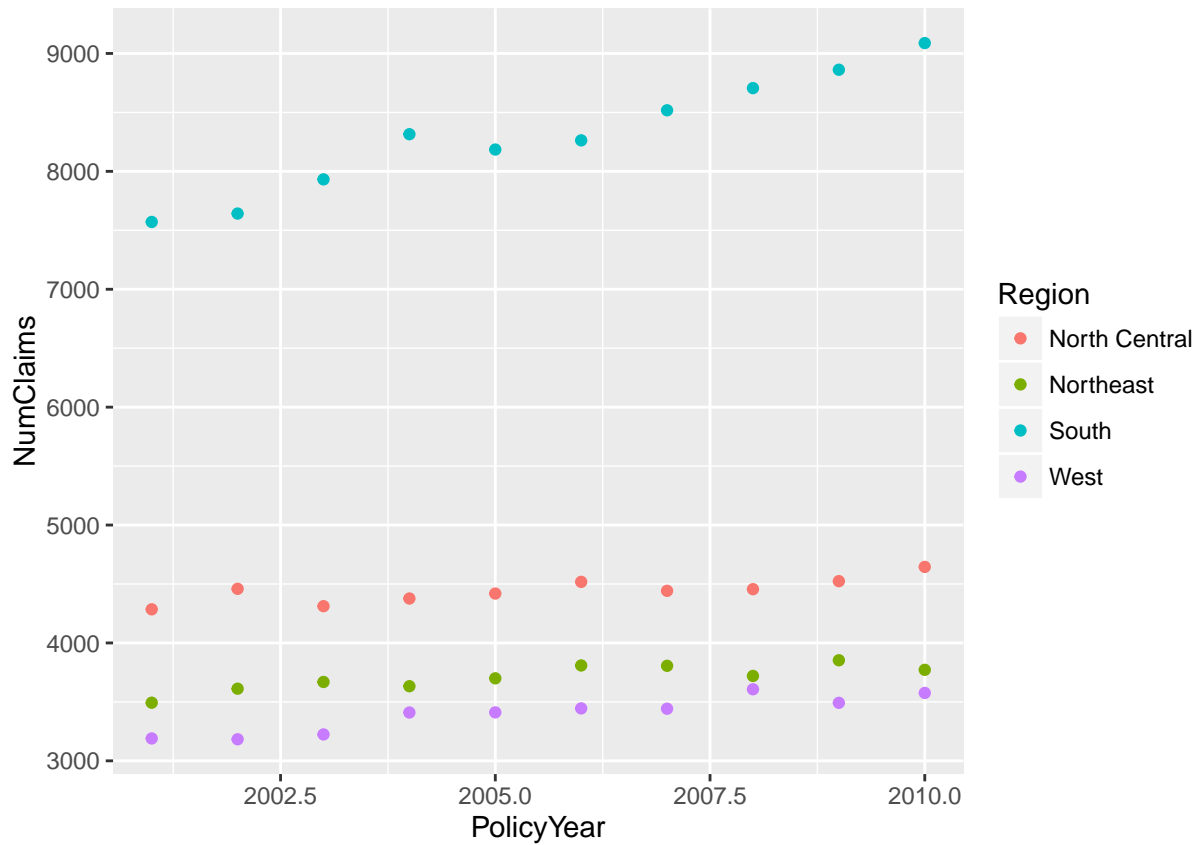
The `geom_*` functions add geometric shapes.

```
p <- basePlot + aes(x = PolicyYear, y = NumClaims, group=Region, color=Region, fill=Region)
p <- p + geom_line()
p
```



Adding layers: points

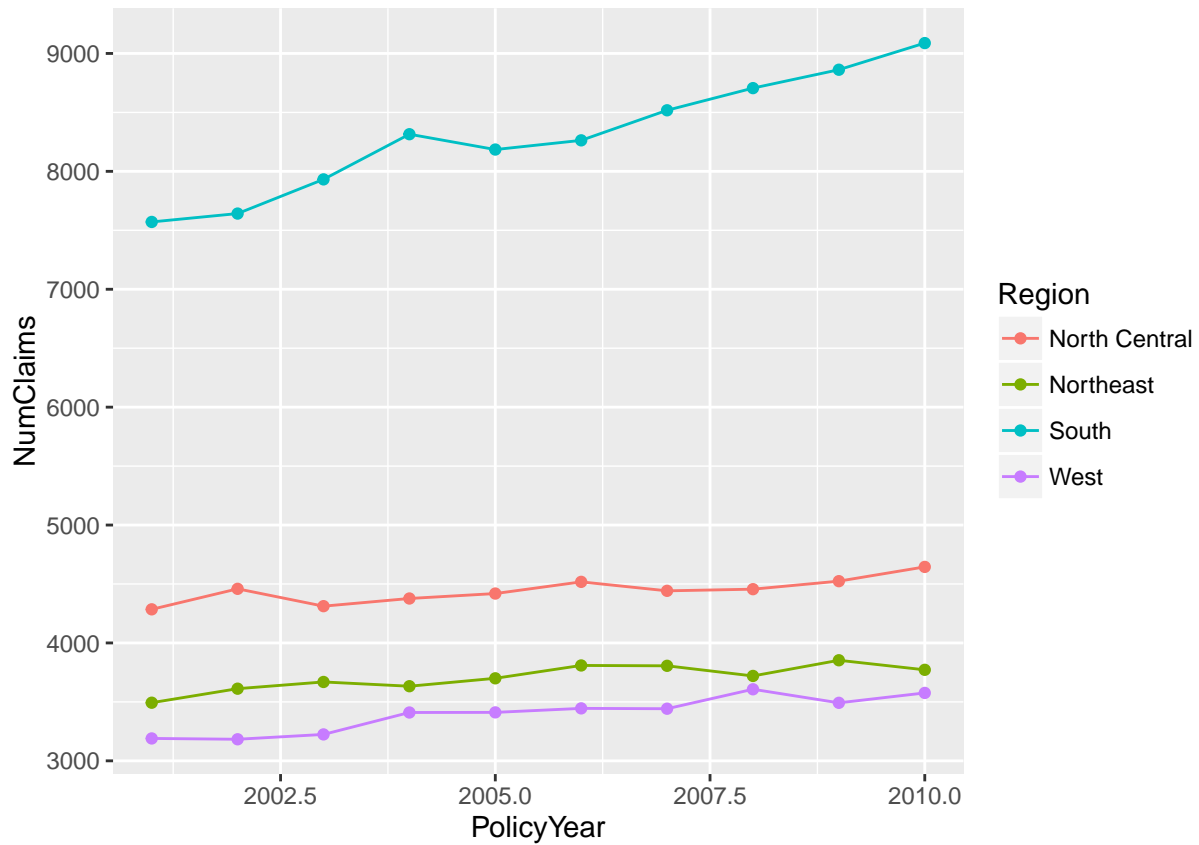
```
p <- basePlot + aes(x = PolicyYear, y = NumClaims, group=Region, color=Region, fill=Region)
p <- p + geom_point()
p
```



Adding layers: lines + points

Nothing wrong with adding two layers

```
p <- basePlot + aes(x = PolicyYear, y = NumClaims, group=Region, color=Region, fill=Region)
p <- p + geom_point() + geom_line()
p
```

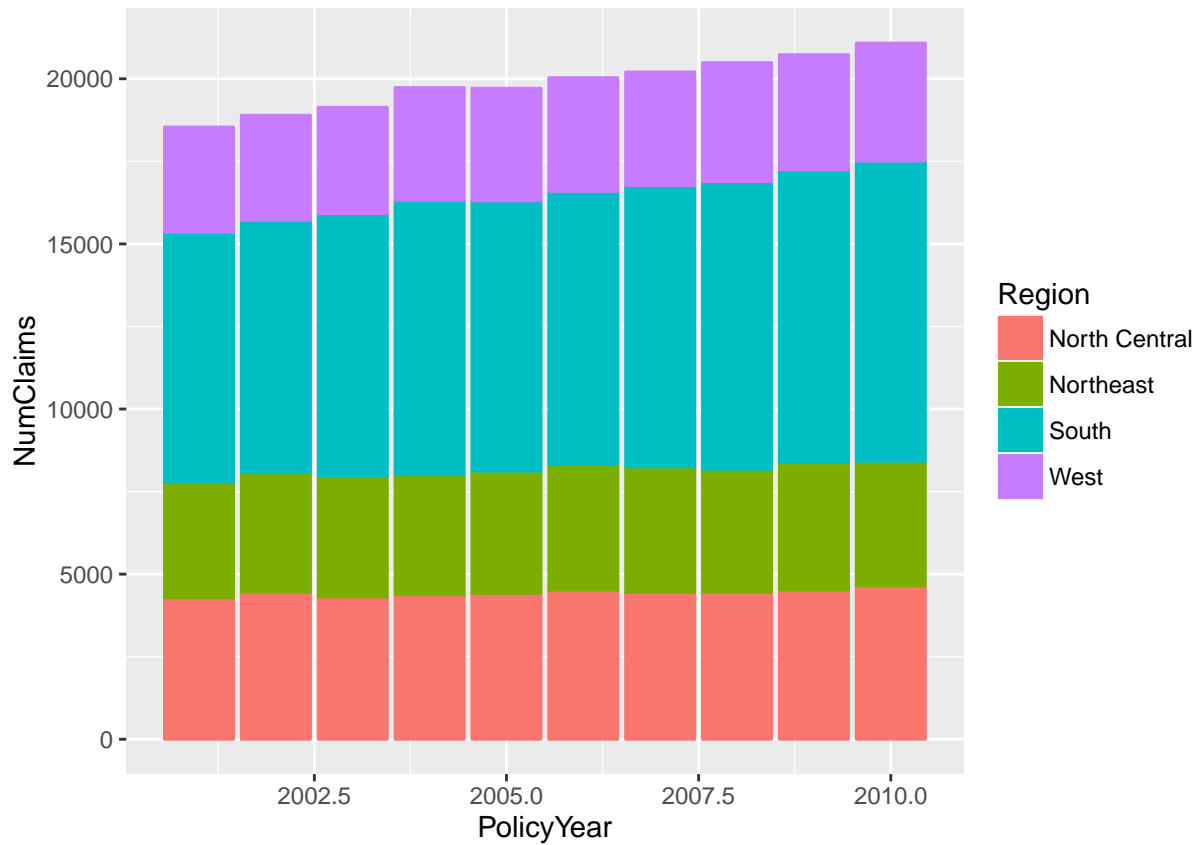


Adding layers: barplot

The heights of the bars commonly represent one of two things:

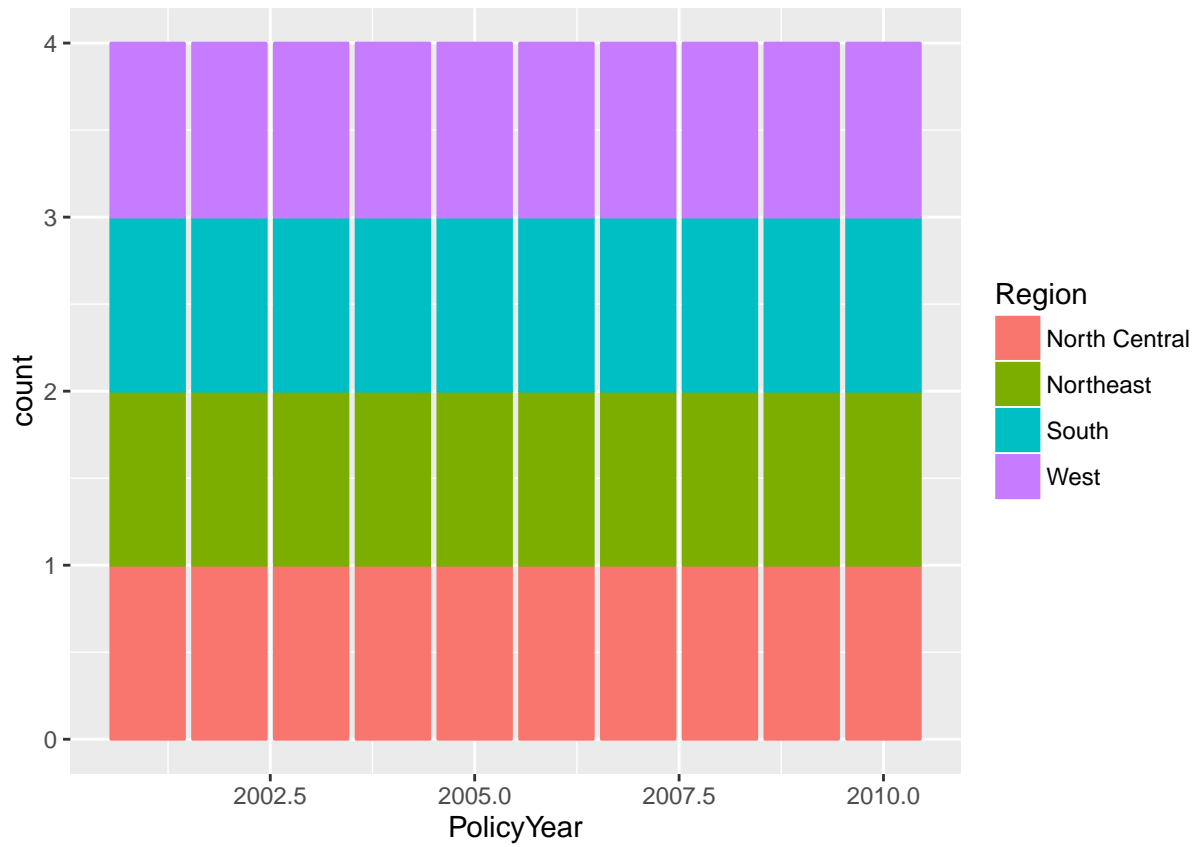
- a count of cases in each group: `stat="identity"`
- the mean values in each group: `stat="bin"` (default)

```
p <- basePlot + aes(x = PolicyYear, y = NumClaims, group=Region, color=Region, fill=Region)
p <- p + geom_bar(stat="identity")
p
```



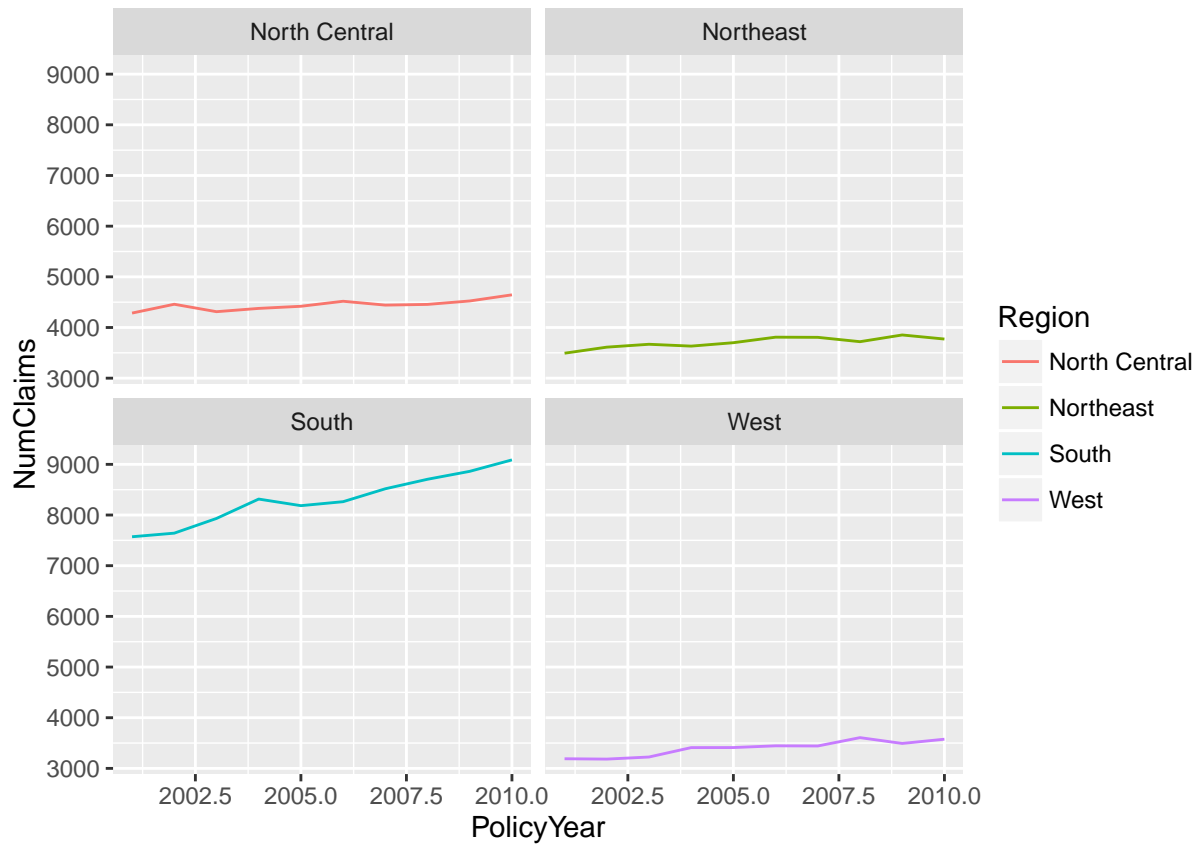
Adding layers: barplot

```
p <- basePlot + aes(x = PolicyYear, group=Region, color=Region, fill=Region)
p <- p + geom_bar()
p
```



Facets

```
p <- basePlot + aes(x = PolicyYear, y = NumClaims, group=Region, color=Region, fill=Region)
p <- p + geom_line() + facet_wrap(~ Region)
p
```

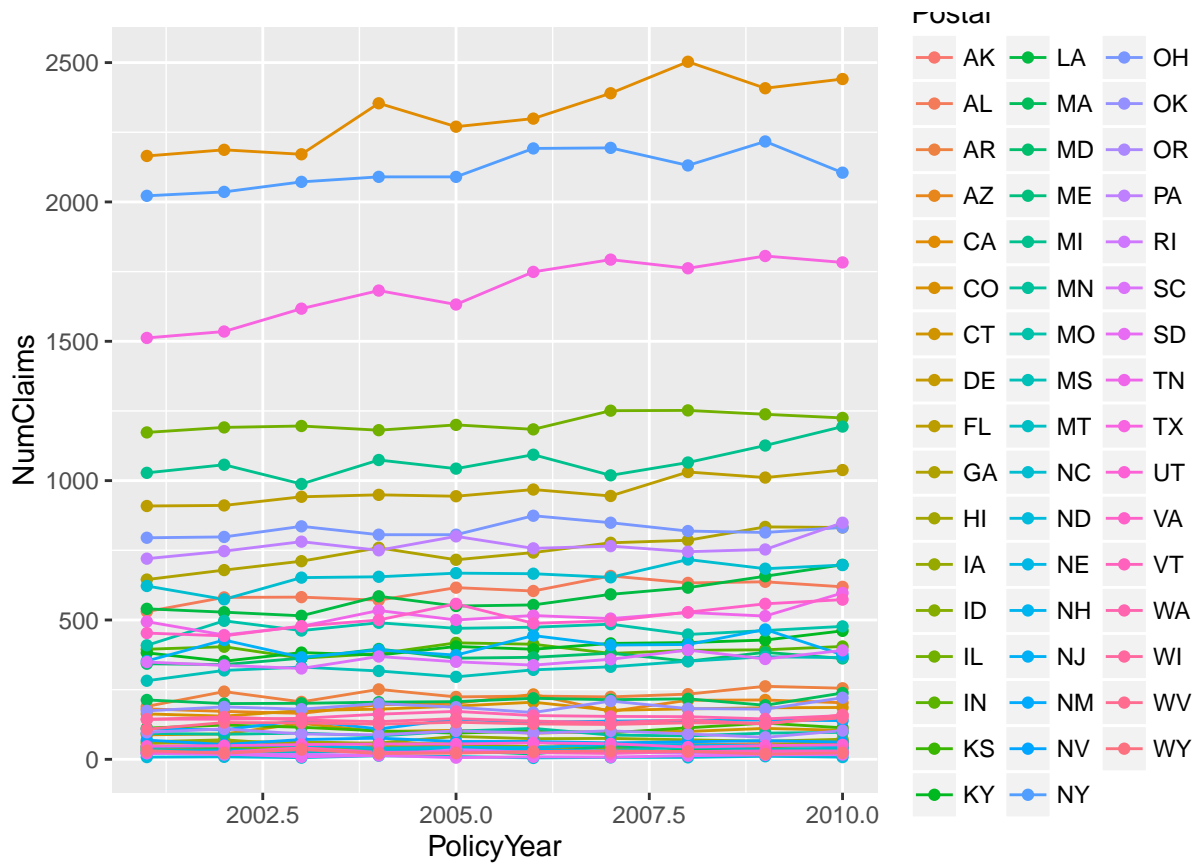


Exercise

```
library(raw)
library(ggplot2)
data(StateExperience)
head(StateExperience)
```

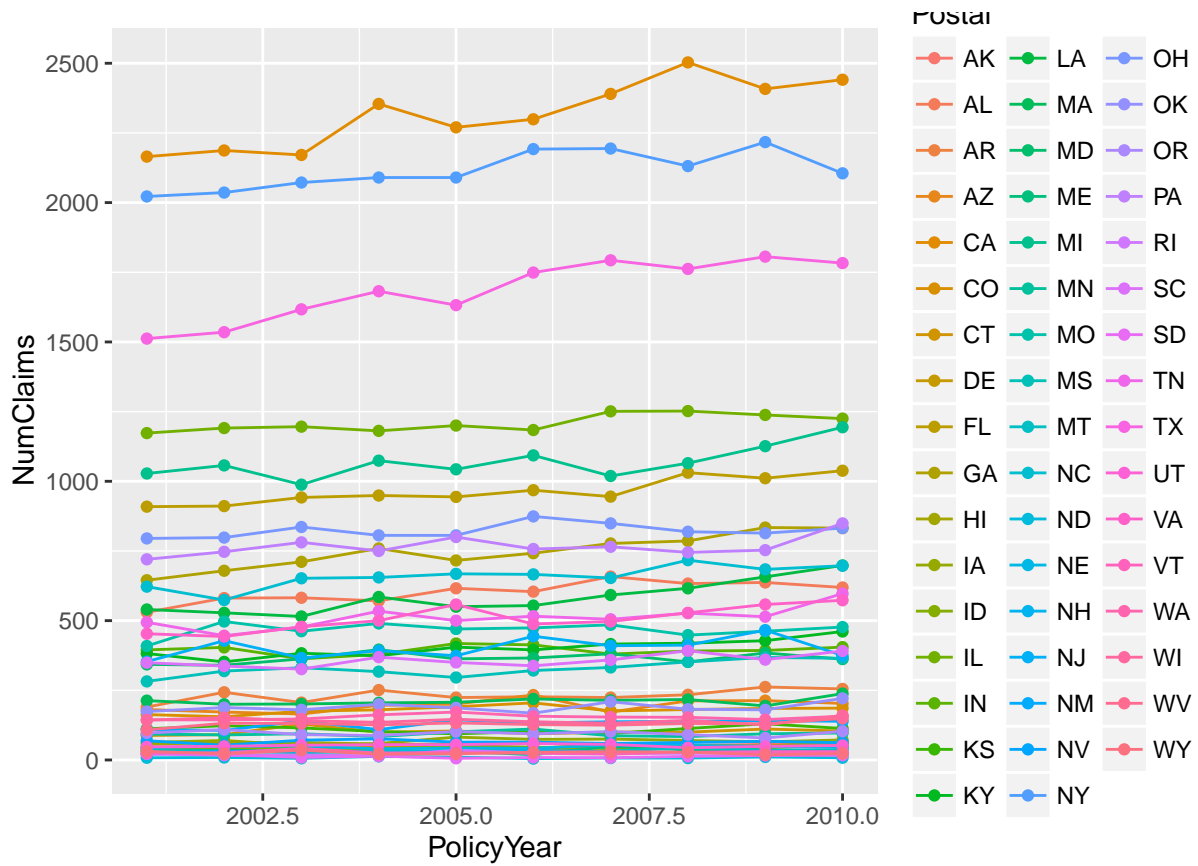
```
##   Region   State Postal PolicyYear NumPolicies NumClaims
## 1  South Alabama   AL      2001         37         530
## 2  South Alabama   AL      2002         38         581
## 3  South Alabama   AL      2003         38         582
## 4  South Alabama   AL      2004         39         571
## 5  South Alabama   AL      2005         40         616
## 6  South Alabama   AL      2006         41         604
```


Lines + Points

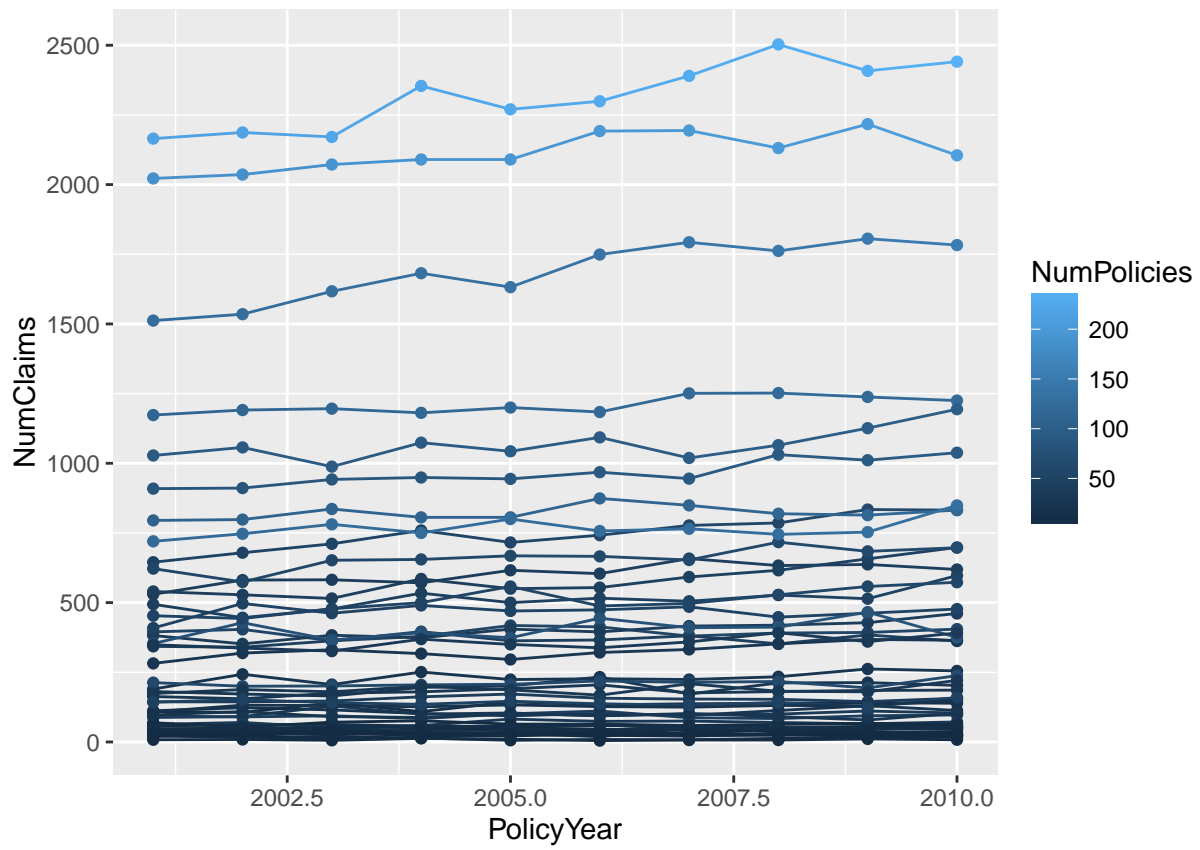


Lines + Points

```
p <- basePlot+ aes(x = PolicyYear, y = NumClaims, group=Postal, color=Postal, fill=Postal)
p <- p + geom_line()+ geom_point()
p
```

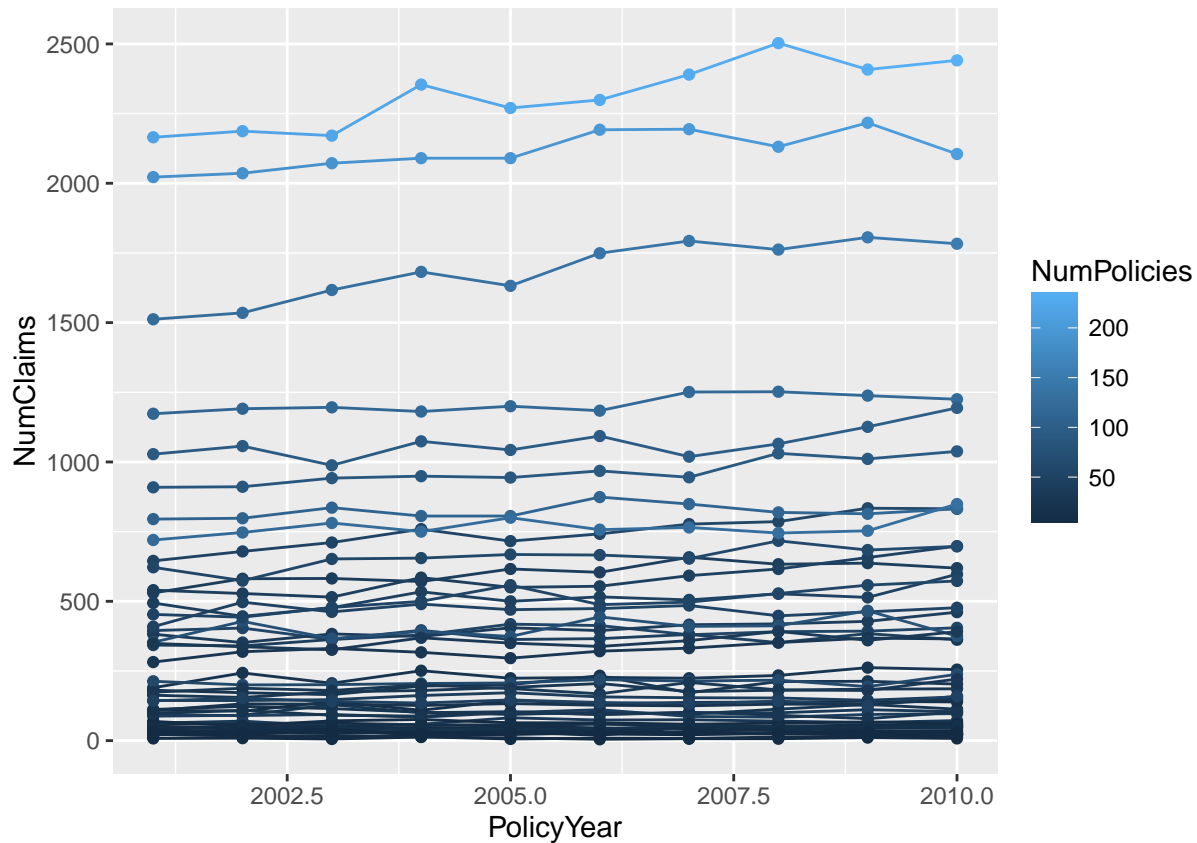


Colors



Colors

```
p <- basePlot+ aes(x = PolicyYear, y = NumClaims, group=Postal, color=NumPolicies)
p <- p + geom_line()+ geom_point()
p
```



Maps and Choropleths

```
library(maps)
crimes <- data.frame(state = tolower(rownames(USArrests)), USArrests)
head(crimes)
```

```
##           state Murder Assault UrbanPop Rape
## Alabama    alabama  13.2   236     58 21.2
## Alaska     alaska   10.0   263     48 44.5
## Arizona    arizona   8.1   294     80 31.0
## Arkansas   arkansas  8.8   190     50 19.5
## California california 9.0   276     91 40.6
## Colorado   colorado  7.9   204     78 38.7
```

```
library(reshape2) # for melt
crimesm <- melt(crimes, id = 1)
head(crimesm)
```

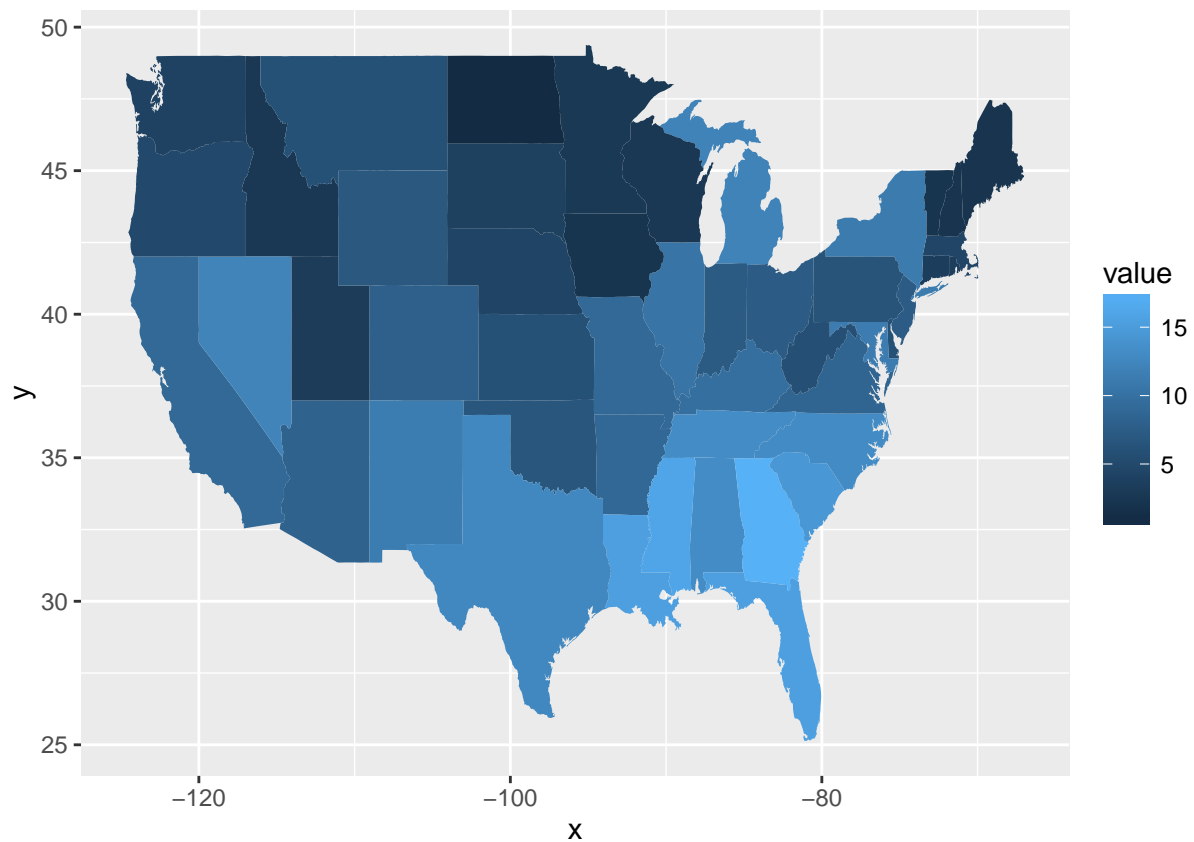
```
##      state variable value
## 1  alabama  Murder   13.2
## 2  alaska   Murder   10.0
## 3  arizona  Murder    8.1
## 4  arkansas Murder    8.8
## 5  california Murder   9.0
## 6  colorado Murder    7.9
```

```
states_map <- map_data("state")
head(states_map)
```

```
##      long      lat group order region subregion
## 1 -87.46201 30.38968     1     1 alabama      <NA>
## 2 -87.48493 30.37249     1     2 alabama      <NA>
## 3 -87.52503 30.37249     1     3 alabama      <NA>
## 4 -87.53076 30.33239     1     4 alabama      <NA>
## 5 -87.57087 30.32665     1     5 alabama      <NA>
## 6 -87.58806 30.32665     1     6 alabama      <NA>
```

Map

```
basePlot <- ggplot(crimesm[crimesm$variable=='Murder',])
p <- basePlot + aes(map_id = state)
p <- p + geom_map(aes(fill = value), map = states_map) + expand_limits(x = states_map$long, y = states_map$lat)
p
```



Facets

```
basePlot <- ggplot(crimesm)
p <- basePlot + aes(map_id = state)
p <- p + geom_map(aes(fill = value), map = states_map) + expand_limits(x = states_map$long, y = states_map$lat)
p
```

