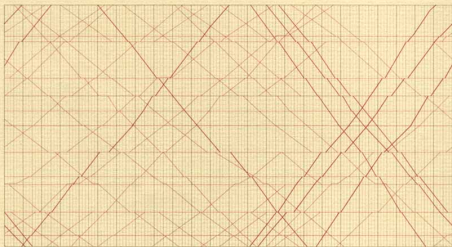


# data visualization

skimming deep waters

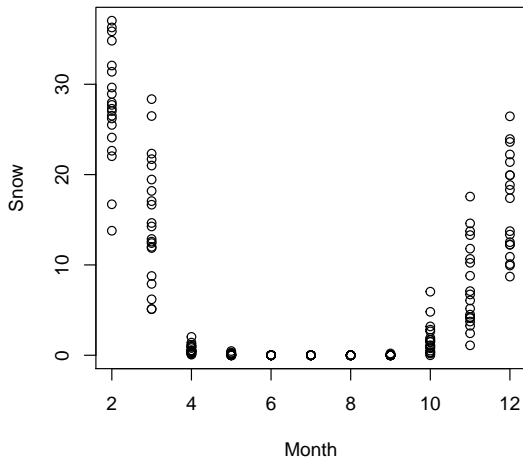




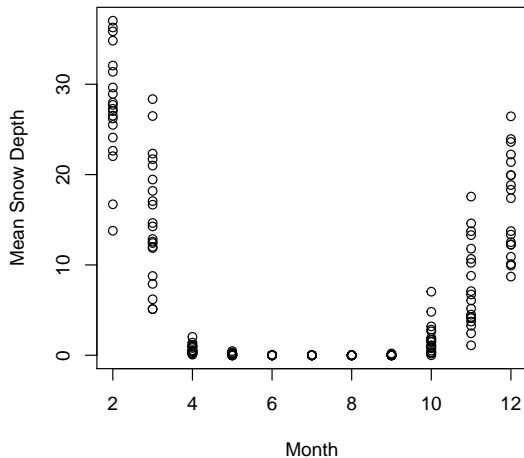
# The Visual Display of Quantitative Information

EDWARD R. TUFTE

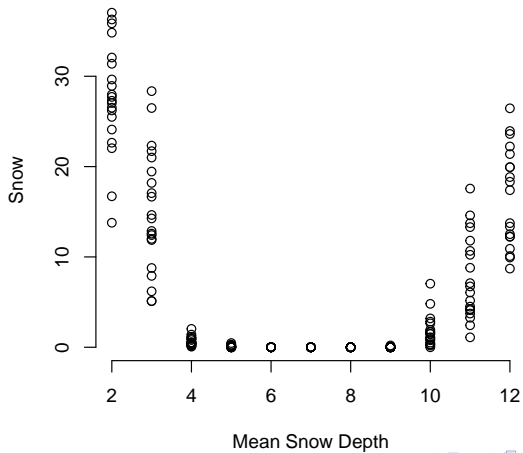
## Show the Data



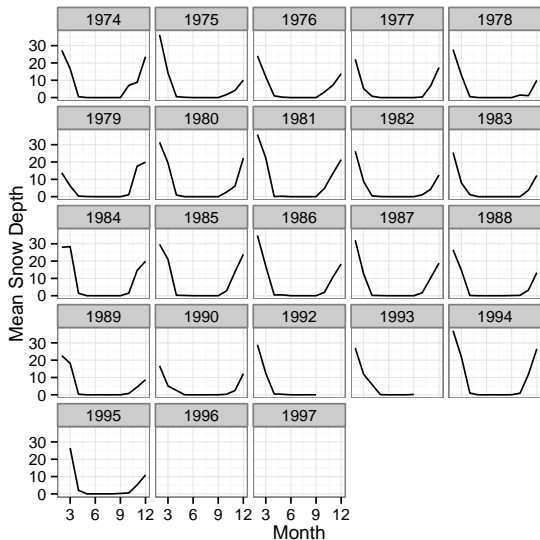
# Minimize Distraction



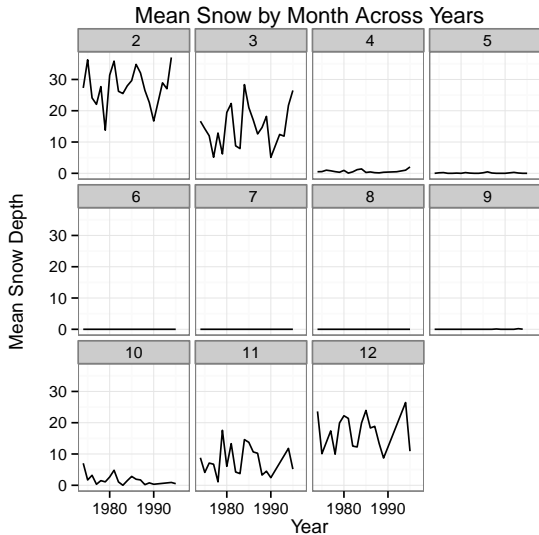
# Minimize Distraction



# Make Big Data Coherent

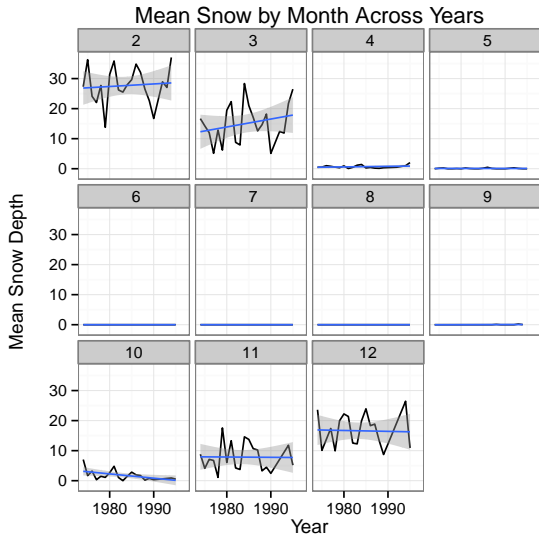


# Reveal Several Levels of Detail





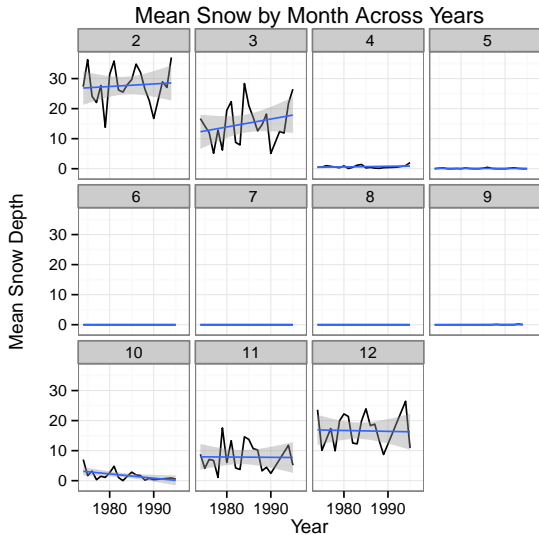
# Be Closely Integrated with Statistics



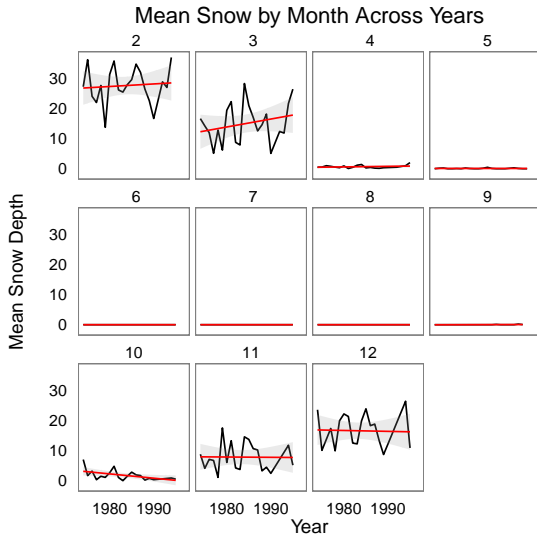
# The Data:Ink Ratio

1. Above all else show data.
2. Maximize the data-ink ratio.
3. Erase non-data-ink.
4. Erase redundant data-ink.
5. Revise and edit

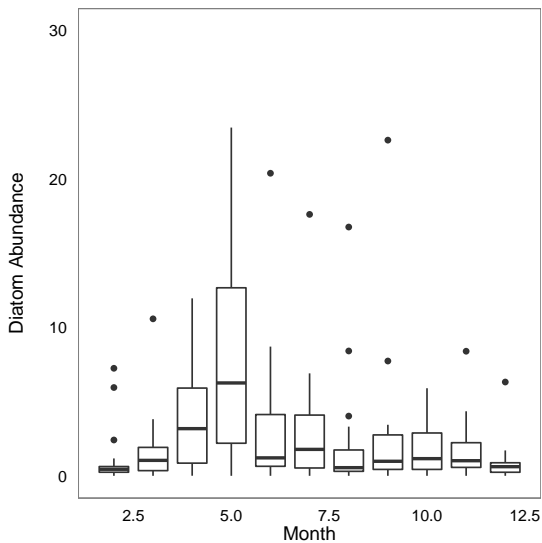
# Minimizing Ink



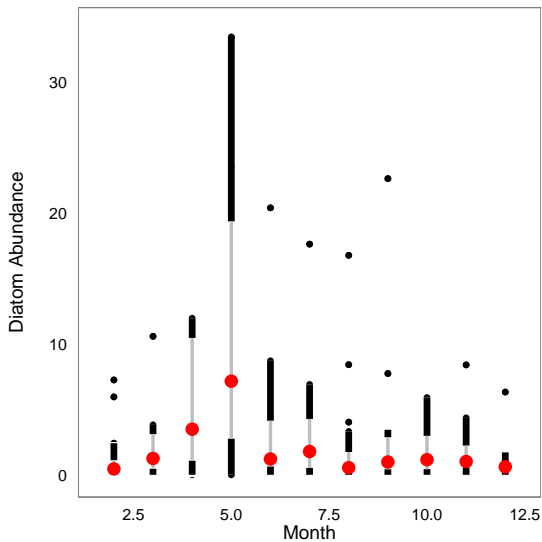
# Minimizing Ink



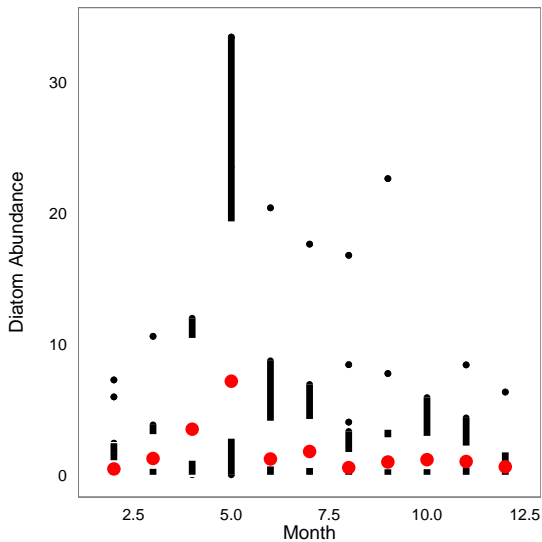
## Extra Ink in Boxes & Lines



## A Cleaner Boxplot



# Pure Tufte Boxplots

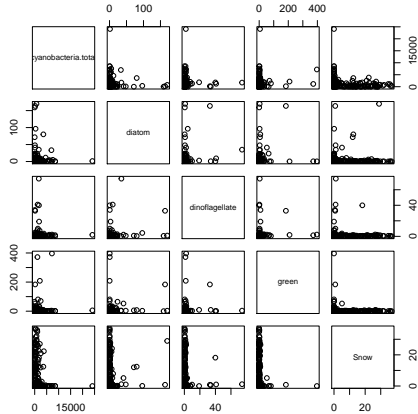


# Basic Plotting in R



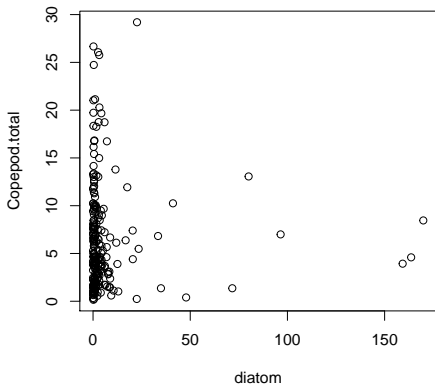
# Visualizing a Lot of the Data

```
pairs(plankton[, 14:18])
```



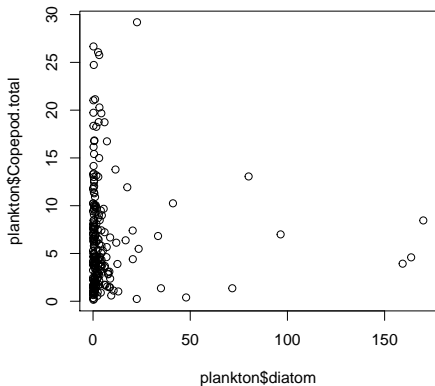
# A Basic Bivariate Plot

```
plot(Copepod.total ~  
      diatom, data = plankton)
```



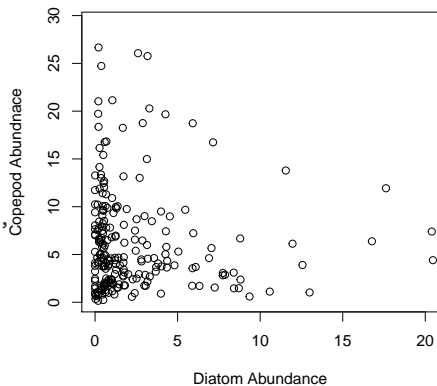
## A Basic Bivariate Plot

```
plot(plankton$diatom,  
     plankton$Copepod.total)
```



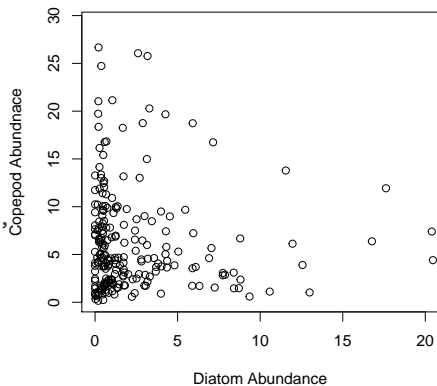
# Adding Axis Labels

```
plot(Copepod.total ~  
      diatom, data = plankton,  
      xlab = "Diatom Abundance",  
      ylab = "Copepod Abundance",  
      xlim = c(0, 20))
```



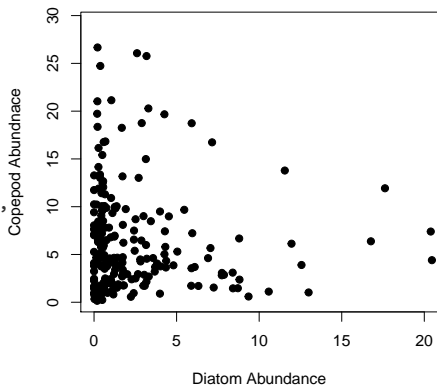
## Adding Axis Limits

```
plot(Copepod.total ~  
      diatom, data = plankton,  
      xlab = "Diatom Abundance",  
      ylab = "Copepod Abundance",  
      xlim = c(0, 20))
```



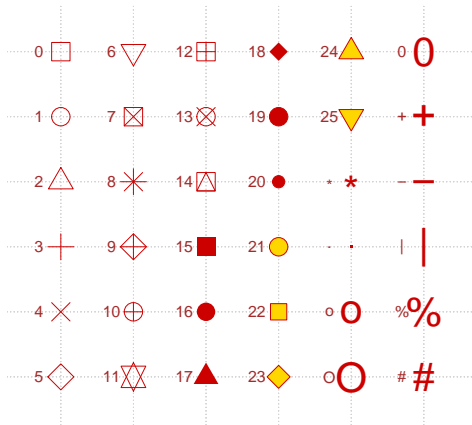
## More Point Shapes

```
plot(Copepod.total ~  
      diatom, data = plankton,  
      xlab = "Diatom Abundance",  
      ylab = "Copepod Abundnace",  
      xlim = c(0, 20),  
      pch = 19)
```



# More Point Shapes

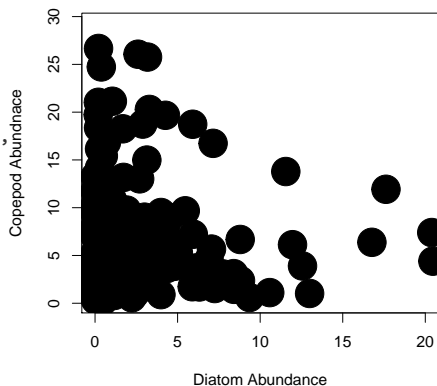
plot symbols : points (... pch = \*, cex = 3)



## cex for Size

```
plot(Copepod.total ~  
      diatom, data = plankton,  
      xlab = "Diatom Abundance",  
      ylab = "Copepod Abundnace",  
      xlim = c(0, 20),  
      pch = 19, cex = 4)
```

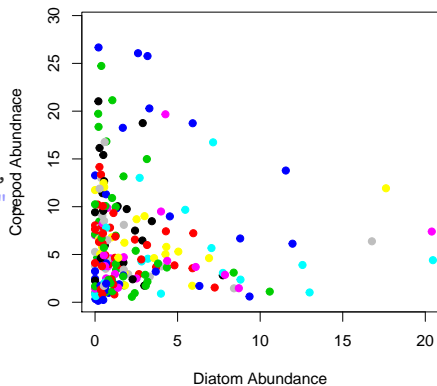
See also `cex.axis`,  
`cex.lab`, and more.





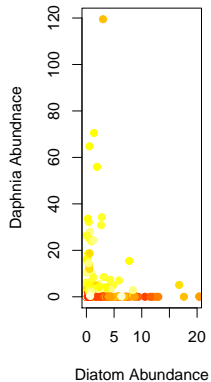
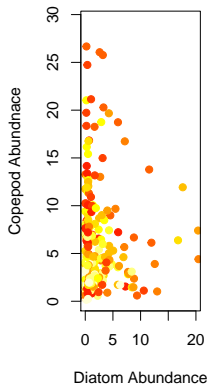
## Add a Little Color

```
plot(Copepod.total ~  
      diatom, data = plankton,  
      xlab = "Diatom Abundance",  
      ylab = "Copepod Abundnace",  
      xlim = c(0, 20),  
      pch = 19, col = Month)
```



## Panels with Par and Mfrow

```
par(mfrow = c(1, 2))
```



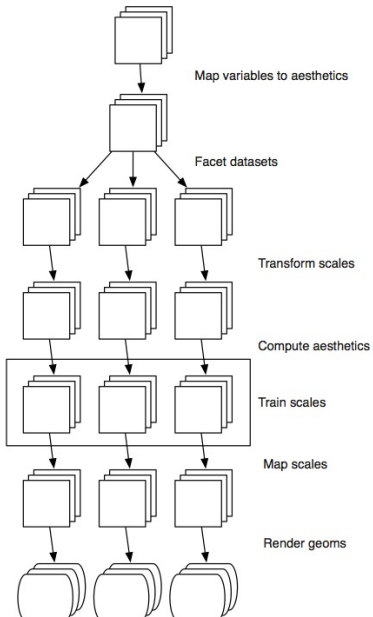
# Lots of Other Functions that For Plots

```
?matplot  
?lines  
?axis  
?title  
?legend  
?points  
?segments
```

So....Explore! Plot with the data, try different par settings, or use some of these functions!

# ggplot2

or how I learned to stop worrying and love <http://had.co.nz/ggplot2> & <http://stackoverflow.com/>



## Start with nothing...

```
p <- ggplot(data = plankton, mapping = aes(x = Month,  
      y = Copepod.total))
```

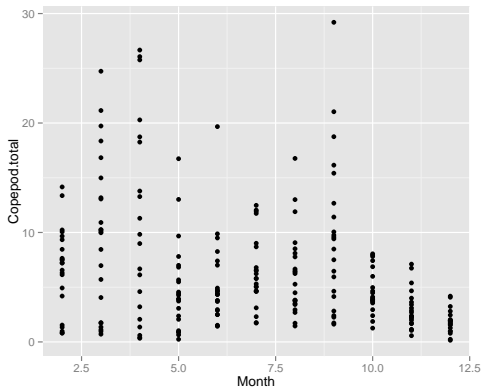
```
p
```

```
## Error: No layers in plot
```

There is no layout specified here for the data.

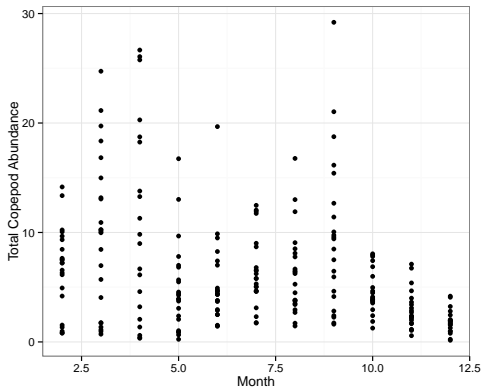
# Add a Layer

```
p <- p + geom_point()  
p
```



# Format with Theme

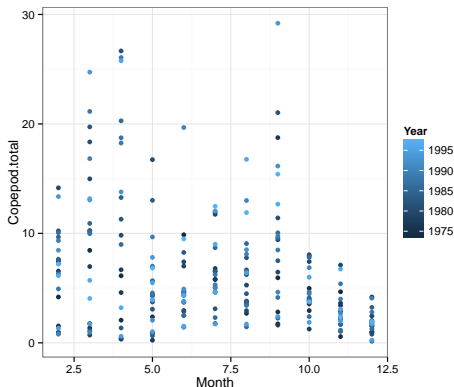
```
p <- p + ylab("Total Copepod Abundance") + theme_bw()  
p
```





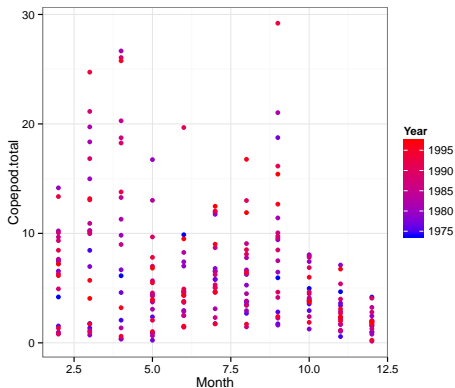
## Map a Variable to Color

```
p2 <- ggplot(data = plankton, aes(x = Month, y = Copepod.total,  
  color = Year)) + geom_point() + theme_bw()  
p2
```



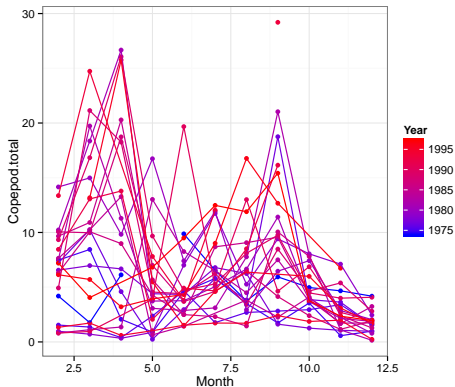
# Set Your Own Scale

```
p2 <- p2 + scale_color_gradient(low = "blue", high = "red")  
p2
```



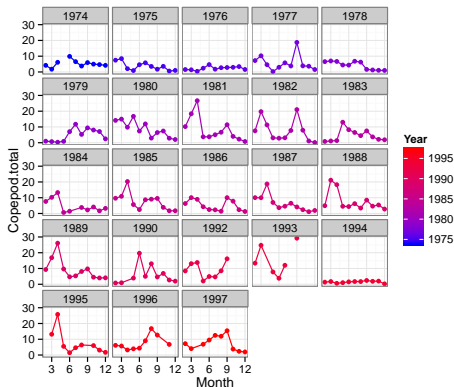
## And Maybe Add Another Layer

```
p2 <- p2 + geom_line(aes(group = Year))  
p2
```



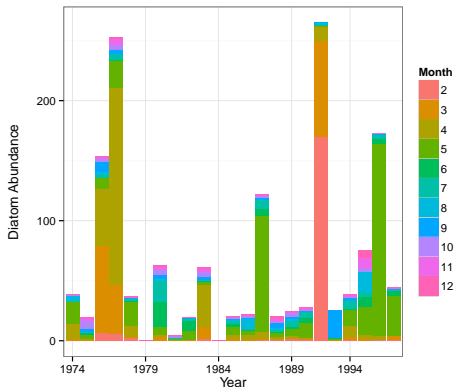
# Facet for Easier Visualization

```
p2 <- p2 + facet_wrap(~Year) + scale_x_continuous(breaks = c(3, 6, 9, 12))  
p2
```



# This All Can Lead to Interesting Visualizations

```
qplot(factor(Year), diatom, geom = "bar", fill = factor(Month),  
  data = plankton) + theme_bw() + xlab("Year") +  
  ylab("Diatom Abundance\n") + scale_fill_discrete(name = "Month") +  
  scale_x_discrete(breaks = seq(1974, 1997, 5))
```



# Lots of Layers to Add to ggplot2 Objects

```
?theme  
?labs  
?xlim  
?facet_grid  
?scale_x_log10  
?geom_histogram  
?geom_ribbon  
?geom_linerange  
?geom_freqpoly
```

So....Explore!

Also, see <http://had.co.nz/ggplot2> for some examples