Making Pretty Pictures with `ggplot2`

Drew Conway

December 3, 2009

Scatter Plots

- ▶ Identifying key actors in #rstats network
- ▶ Tipping: Women vs. Men

Time-Series

- ▶ Closing Price of Four Major European Stock Indices 1991-1998
- ▶ Effect of Seatbelt Law on Automobile Fatalities in UK

Other Examples and Online Resources

- ▶ Learning R Blog
- ▶ Hadley's Resources
- ▶ REvolution Computing Blog

Introduction
Scatter Plots
Time Series Plots
Other Examples and Online Resources

Key Actor Analysis
Tipping by Sex

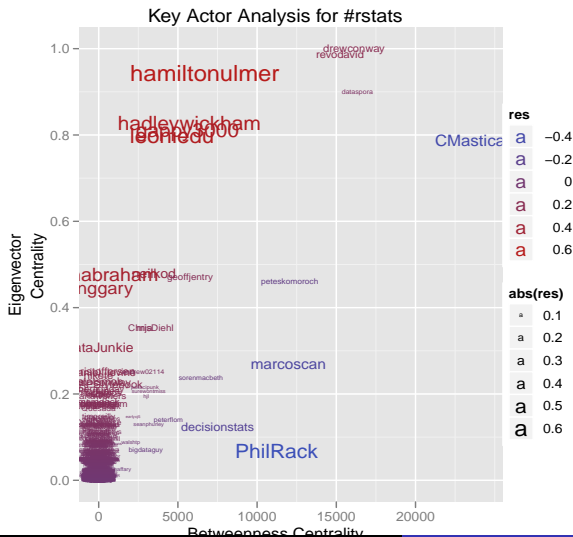## Identifying key actors in #rstats network

Last month I presented the SNA in R talk to the Bay Area R Meetup

- ▶ As part of this presentation I generated the online social network of people on Twitter discussing R by mining the `#rstats` hash-tag
- ▶ As before, I used `ggplot2` to generate a plot that highlights key actors in the network

```
> # Load ggplot2 library
> library(ggplot2)
> # Load in centrality data
> rstats.cent<-read.csv("rstats_centrality.csv")
> # Calculate residuals for simple OLS of eig~bet
> rstats.res<-as.vector(lm(eig~bet,data=rstats.cent)$residuals)
> rstats.cent<-transform(rstats.cent,res=rstats.res)
> # Create plot
> cent.plot<-ggplot(rstats.cent,aes(x=bet,y=eig,label=name,colour=res,
+     size=abs(res)))+xlab("Betweenness Centrality")+ylab("Eigenvector
+     Centrality")
> cent.plot<-cent.plot+geom_text()+opts(title="Key Actor Analysis for #rstats")
```

Introduction
Scatter Plots
Time Series Plots
Other Examples and Online Resources

Key Actor Analysis
Tipping by Sex

Key Actor Plot of #rstats Social Network

This plot will look familiar to those that attended my SNA in R talk in August!

Introduction
Scatter Plots
Time Series Plots
Other Examples and Online Resources
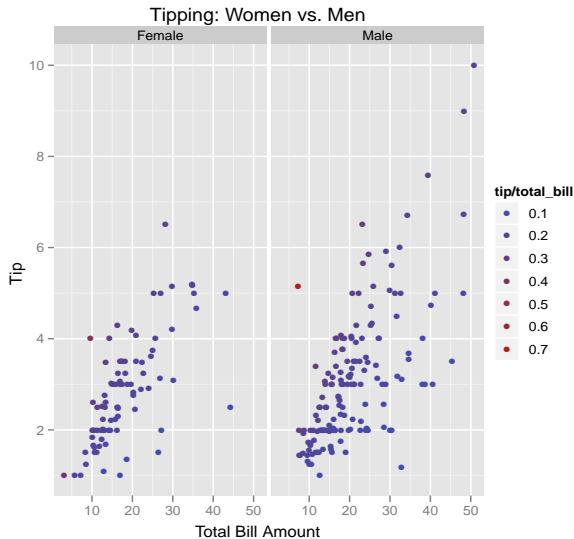
Key Actor Analysis
Tipping by Sex

Tipping: Women vs. Men

Using the pre-loaded data `tips` we will create a simple comparative scatter plot of total bill amount vs. the subsequent tip paid by both women and men.

```
> # Create plot
> tip.bysex<-ggplot(tips,aes(total_bill))+
+     geom_point(aes(y=tip,colour=tip/total_bill))+xlab("Total Bill Amount")+
+     ylab("Tip")+facet_wrap(~sex,ncol=2)
> tip.bysex<-tip.bysex+opts(title="Tipping: Women vs. Men")
```

▶ By using the `facet_wrap()` we will have two plots in a single chart; each corresponding to the tipping data for women and men respectively

▶ By setting `colour=tip/total_bill` the plot will highlight which sex contains more "outlier tippers"

Introduction
Scatter Plots
Time Series Plots
Other Examples and Online Resources

Key Actor Analysis
Tipping by Sex

Plot of Tipping Trends by Sex

Introduction
Scatter Plots
Time Series Plots
Other Examples and Online Resources

Major EU Stock Indices
Driver Fatalities in the U.K. and Seatbelt Law

Plotting the Daily Closing Price of Four Major EU Stock Indices

Using the base `datasets` package we will plot the daily closing price of the DAX, SMI, CAC and FTSE from 1991 through 1998
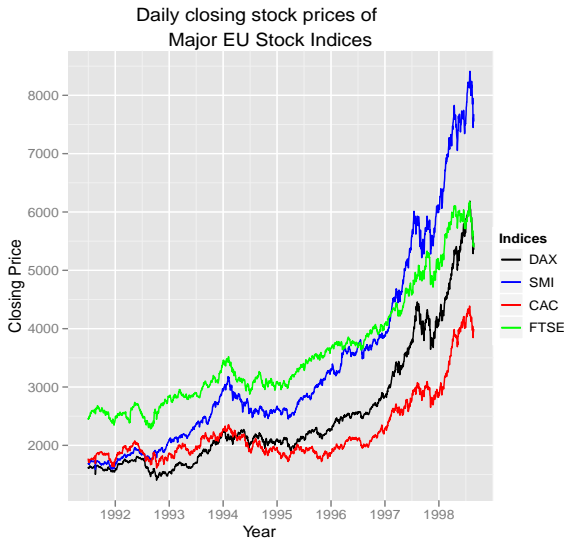
▶ This data set is a `ts` object, so we must first convert it to a data frame with a time vector

```
> # Create data frame and add time vector
> eu<-data.frame(EuStockMarkets)
> eu<-transform(eu,time=time(EuStockMarkets))
> # Create base plot object
> eu.plot<-ggplot(eu,aes(time))
```

▶ Now add the stock price data and appropriate legend

```
> # Add stock price data
> eu.plot<-eu.plot+geom_line(aes(y=DAX,colour="DAX"))+
+     geom_line(aes(y=SMI,colour="SMI"))
> eu.plot<-eu.plot+geom_line(aes(y=CAC,color="CAC"))+
+     geom_line(aes(y=FTSE,colour="FTSE"))
> # Create title and legend
> eu.plot<-eu.plot+opts(title="Daily closing stock prices of
+     Major EU Stock Indices")+xlab("Year")+ylab("Closing Price")
> eu.plot<-eu.plot+scale_colour_manual(name="Indices",
+     values=c("DAX"="black","SMI"="blue","CAC"="red","FTSE"="green"))
```

Introduction
Scatter Plots
Time Series Plots
Other Examples and Online Resources

Major EU Stock Indices
Driver Fatalities in the U.K. and Seatbelt Law

Plot of Closing Price for Major EU Stock Indices

Daily closing stock prices of Major EU Stock Indices

Introduction
Scatter Plots
Time Series Plots
Other Examples and Online Resources

Major EU Stock Indices
Driver Fatalities in the U.K. and Seatbelt Law

What Impact did Seatbelt Laws have on Driver Fatalities in the UK?

Using the `Seatbelts` dataset we will attempt to visualize the impact of seatbelt legislation on driver deaths in the UK

- ▶ The data contains monthly fatality data from 1969-1984 with a dummy variable for the month/year the seatbelt law went into effect

- ▶ Automobile accidents will likely have a seasonal component, so we will also want to examine the data on a month-by-basis

```
> # Create data frame from ts object and add year/month data
> seatbelts.data<-as.data.frame(Seatbelts)
> seatbelts.data<-transform(seatbelts.data,year=floor(as.vector(time(Seatbelts))
+     month=(rep(1:12,nrow(seatbelts.data/12))))
> # Create a plot that shows seasonality of accidents as well as effect of law
> driver.plot<-ggplot(seatbelts.data,aes(year,DriversKilled,
+     colour=law))+geom_line()+xlab("Year")+ylab("Drivers
+     Killed")+facet_wrap(~month,ncol=6)
> driver.plot<-driver.plot+opts(title="Drivers Fatalities in U.K.
+     by Month from 1969-1984")
```

Introduction
Scatter Plots
Time Series Plots
Other Examples and Online Resources

Major EU Stock Indices
Driver Fatalities in the U.K. and Seatbelt Law

Drivers Fatalities in U.K. by Month from 1969-1984
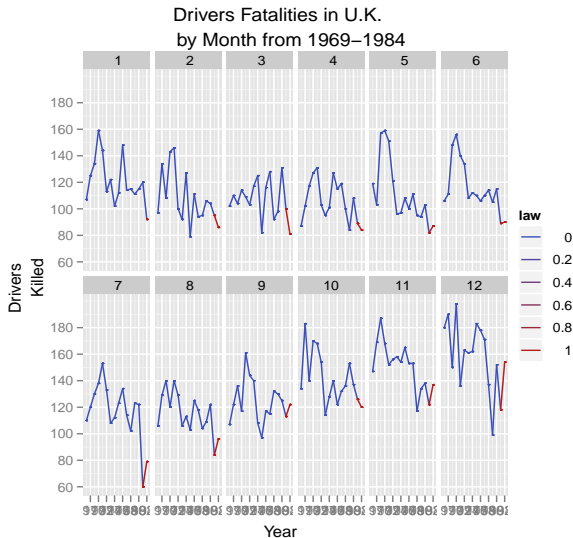
Additional Resources

Learning R Blog

http://learnr.wordpress.com/

An outstanding resource for examples of how to use ggplot2 for several plotting tasks

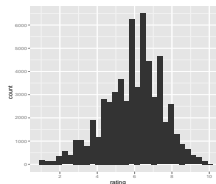- ▶ Some of the most useful posts are the many side-by-side comparisons of lattice and ggplot2

Hadley's Resources

http://had.co.nz/ggplot2/
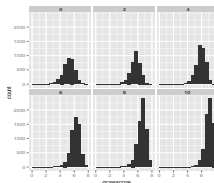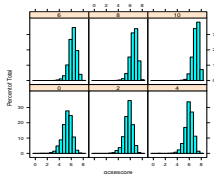
The developer has **numerous** resources and examples on his website

- ▶ The reference manual

  ```
  > data(movies)
  > m <- ggplot(movies, aes(x = rating))
  > m <- m + geom_histogram()
  ```
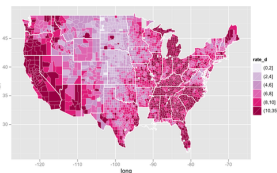

- ▶ The ggplot2 book
  - ▶ Mastering the grammar of graphics
  - ▶ Building up a plot layer by layer
  - ▶ A toolbox of techniques
  - ▶ Scales, axes and legends
  - ▶ Positioning
  - ▶ Manipulating data

REvolution Computing Blog

http://blog.revolution-computing.com/

Choropleth Challenge

- ▶ Plotting the US unemployment data on a color-coded map

Promoting open government in Brazil

- ▶ DC R hacker Eduardo Leoni uses ggplot2 at CongressoAberto.com.br