

Lecture 5: Using ggplot2 for data visualization

LSA 2013, LI539
Mixed Effect Models

Judith Degen



Brain and Cognitive Sciences
University of Rochester

July 12, 2013

What is ggplot?

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- an R package for producing statistical graphics
- a language, based on the Grammar of Graphics (g , a), for describing and creating plots
- plots that can be built up iteratively and edited later

The grammar of graphics

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

What is a graphic?

A mapping from **data** to **aesthetic attributes** (e.g., color, shape, size) of **geometric objects** (e.g., points, lines, bars).

- may contain statistical transformations of the data
- drawn on a specific coordinate system
- can use faceting for generating same plot for different subsets of the dataset

fixme you could insert a graphic here showing this process

Components of a graphic/plot

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- **data** and set of **aesthetic mappings** describing how to map variables in data to aesthetic attributes
- geometric objects (**geoms**) are what you see: points, lines, bars
- statistical transformations (**stats**) summarise data (e.g., binning, counting observations) → optional
- **scales** map values in data space to values in aesthetic space
- **coordinate system** describes how data coordinates are mapped to graphic plane
- **faceting** specification describes how to break data into subsets and return an individual plot for each subset (like conditioning)

Today

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- plot types:
 - histograms
 - density functions
 - bar plots
 - scatterplots
- control plot appearance using
 - scales
 - theme
- error bars

Getting started

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- open file ggplot_exercise.R

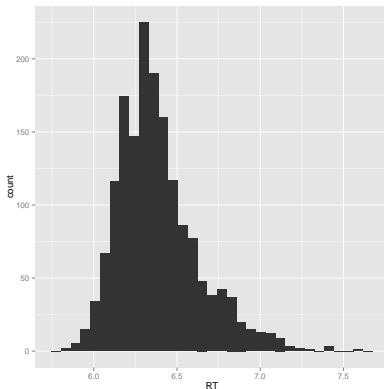
```
library(ggplot2)
library(languageR)
library(Hmisc)
data(lexdec)

# define some functions
se <- function(x)
{
  y <- x[!is.na(x)] # remove the missing values, if any
  sqrt(var(as.vector(y))/length(y))
}
```

First step in understanding your data: histograms

- two parts to every ggplot graphic:
 - `ggplot()` call with `data.frame` and aesthetic mapping argument
 - add graphic layers iteratively

```
# Create a histogram of RTs in the lexdec dataset  
ggplot(lexdec, aes(x=RT)) +  
  geom_histogram()
```



More histograms

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

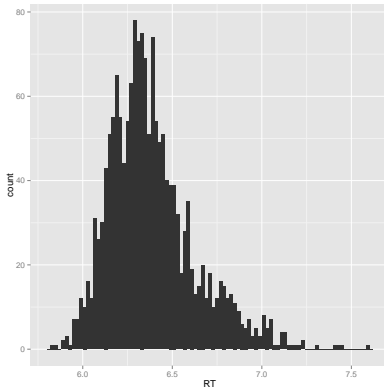
Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- try to change the binwidth



More histograms

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

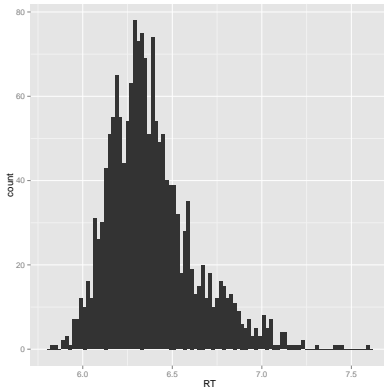
Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

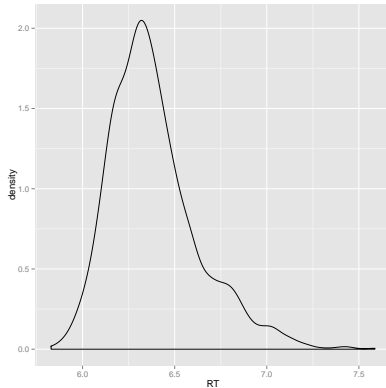
- try to change the binwidth



```
ggplot(lexdec, aes(x=RT)) +  
  geom_histogram(binwidth=0.02)
```

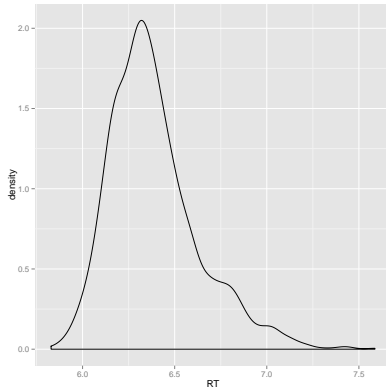
Density plot

- plot density function instead of histogram



Density plot

- plot density function instead of histogram



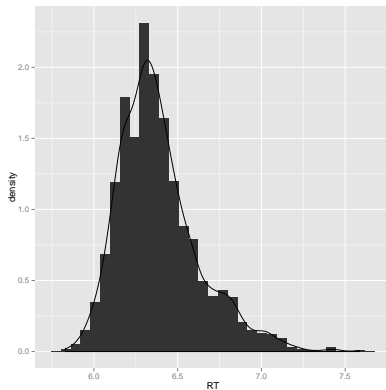
```
ggplot(lexdec, aes(x=RT)) +  
  geom_density()
```

Histogram with overlaid density function

LI539
Mixed
Effect
Models

Degen

- plot density function over histogram



What is
ggplot?

Histograms
and
densities

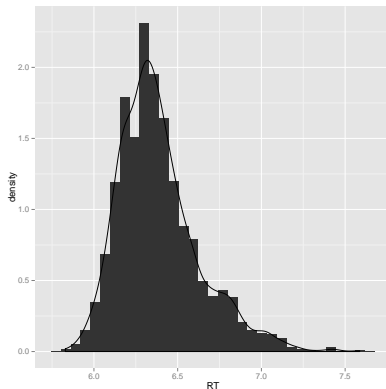
Bar plots

Scatterplots

Manipulating
plot
appearance

Histogram with overlaid density function

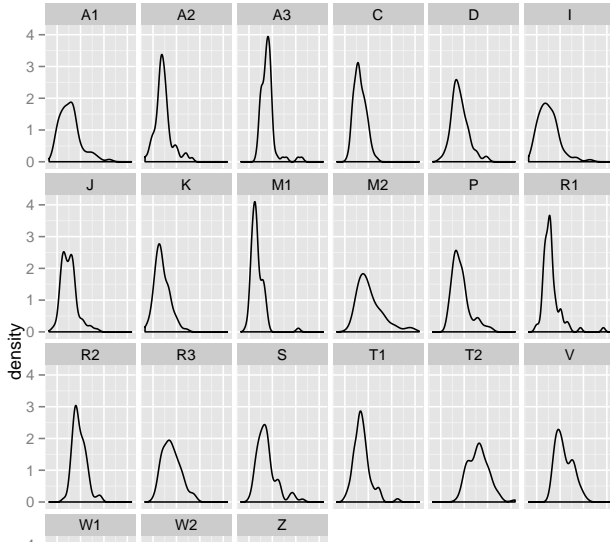
- plot density function over histogram



```
ggplot(lexdec, aes(x=RT)) +  
  geom_histogram(aes(y = ..density..)) +  
  geom_density()
```

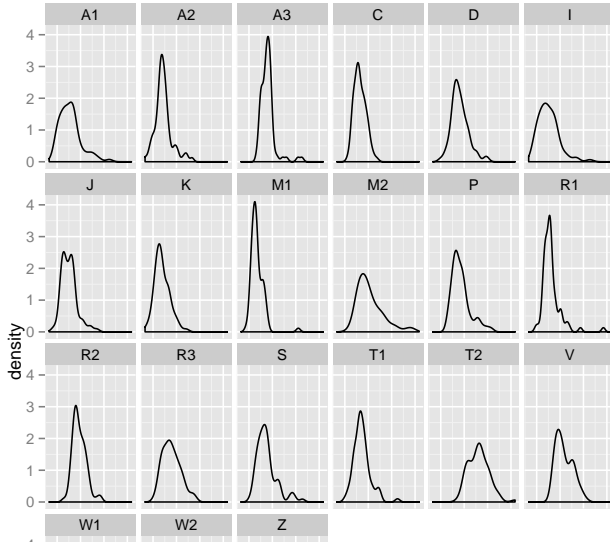
Using facets

- do all subjects have a similar RT distribution?
- create an RT histogram for each subject using facets



Using facets

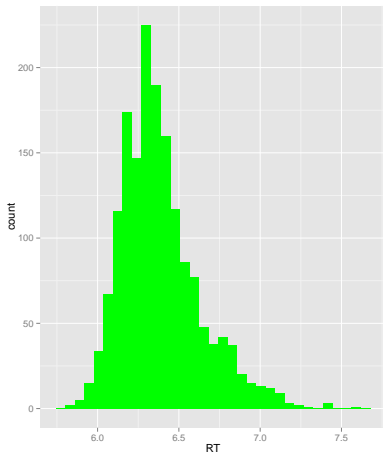
- do all subjects have a similar RT distribution?
- create an RT histogram for each subject using facets



Color - setting vs. mapping

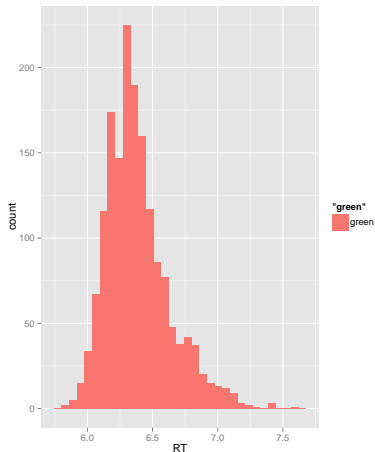
- set color to a constant value

```
ggplot(lexdec, aes(x=RT)) +  
  geom_histogram(fill="green")
```



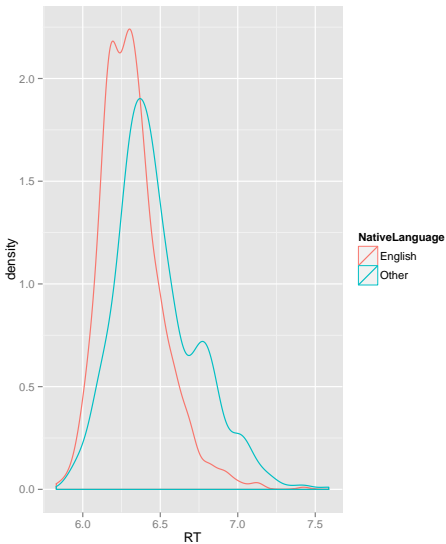
- map color to a variable in aes()

```
ggplot(lexdec, aes(x=RT, fill="green"))  
  geom_histogram()
```



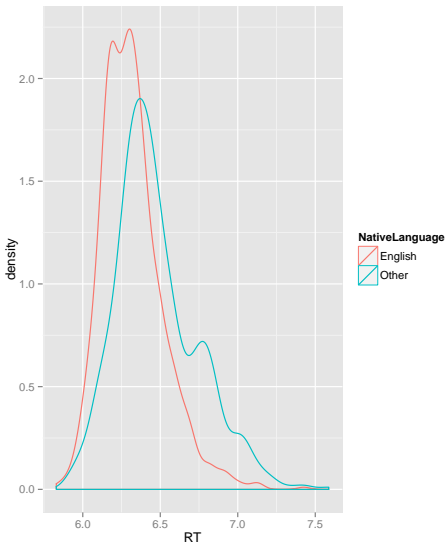
Mapping color to a variable

- plot density functions with different colors for different native languages



Mapping color to a variable

- plot density functions with different colors for different native languages



Bar plots

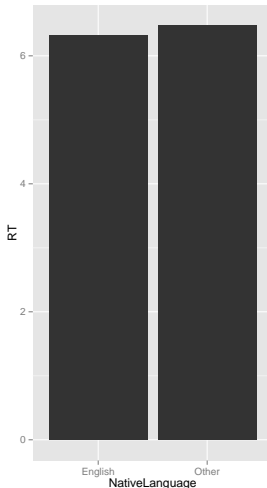
- make a bar plot of mean RTs by native language
- second step: plot.

- first step: aggregate

```
agr = aggregate(RT ~ NativeLanguage,  
                lexdec, mean)
```

```
agr
```

```
##   NativeLanguage    RT  
## 1      English 6.318  
## 2      Other 6.474
```



Bar plots

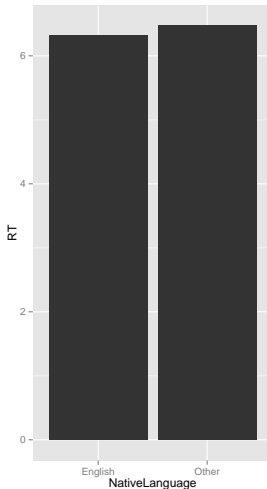
- make a bar plot of mean RTs by native language
- second step: plot.

- first step: aggregate

```
agr = aggregate(RT ~ NativeLanguage,  
                lexdec, mean)
```

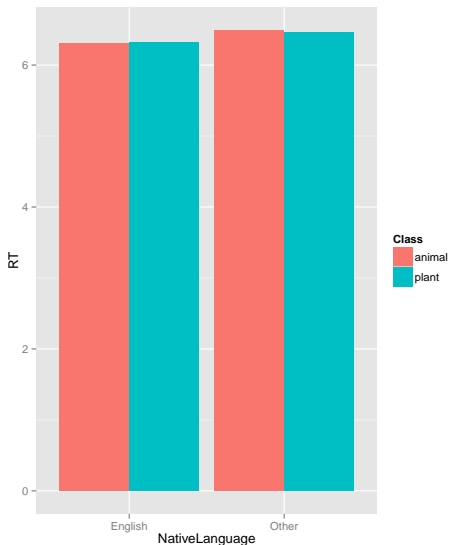
```
agr
```

```
##   NativeLanguage    RT  
## 1      English 6.318  
## 2      Other 6.474
```



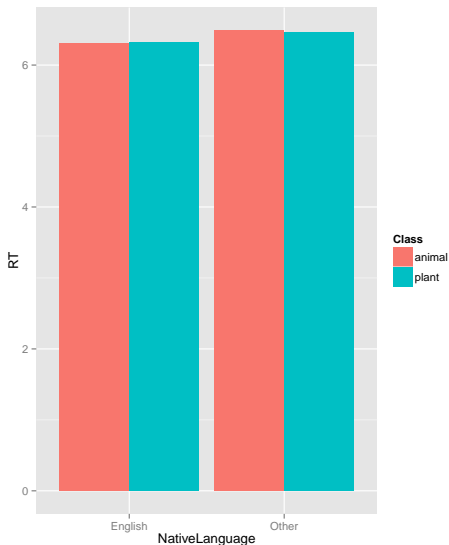
Bar plot with colors

- create same bar plot as before, but with different colors representing different word classes



Bar plot with colors

- create same bar plot as before, but with different colors representing different word classes



LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

Interpretation of plot

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- are differences between conditions real?
- two ways to find out:

Interpretation of plot

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- are differences between conditions real?
- two ways to find out:
 - stats
 - add error bars to plot

Interpretation of plot

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- are differences between conditions real?
- two ways to find out:
 - stats
 - add error bars to plot
- what should error bars represent?

Interpretation of plot

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

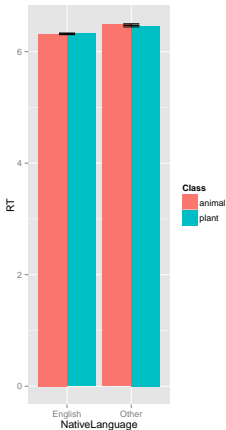
Scatterplots

Manipulating
plot
appearance

- are differences between conditions real?
- two ways to find out:
 - stats
 - add error bars to plot
- what should error bars represent?
 - 1 SE of mean (68% confidence intervals)
 - 95% confidence intervals

Bar plots with error bars

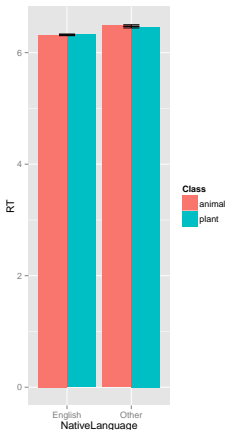
- error bars: 1 SE



```
agr = aggregate(RT ~ NativeLanguage +  
                Class, lexdec, mean)  
agr$SE = aggregate(RT ~ NativeLanguage  
                  + Class, lexdec, se)$RT  
agr$YMin = agr$RT - agr$SE  
agr$YMax = agr$RT + agr$SE
```

Bar plots with error bars

- error bars: 1 SE



```
agr = aggregate(RT ~ NativeLanguage +  
                Class, lexdec, mean)  
agr$SE = aggregate(RT ~ NativeLanguage  
                  + Class, lexdec, se)$RT  
agr$YMin = agr$RT - agr$SE  
agr$YMax = agr$RT + agr$SE
```

```
ggplot(agr, aes(x=NativeLanguage, y=RT, fill=Class)) +  
  geom_bar(stat="identity", position="dodge") +  
  geom_errorbar(aes(ymin=YMin, ymax=YMax),width=0.25)
```

Let's improve this plot.

Adjust error bar position

L1539
Mixed
Effect
Models

Degen

What is
ggplot?

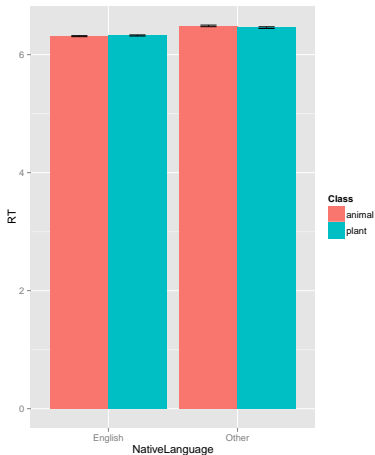
Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

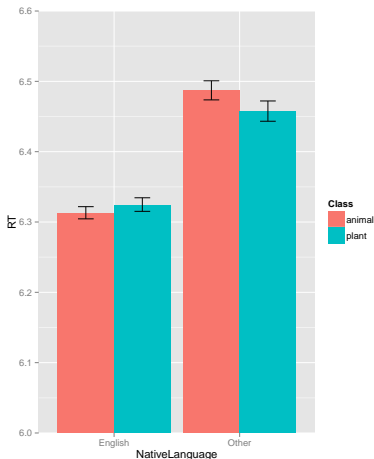
```
dodge = position_dodge(.9)
ggplot(agr, aes(x=NativeLanguage, y=RT, fill=Class)) +
  geom_bar(stat="identity", position=dodge) +
  geom_errorbar(aes(ymin=YMin, ymax=YMax), width=0.25, position=dodge)
```



Zoom into plot by restricting coordinate system

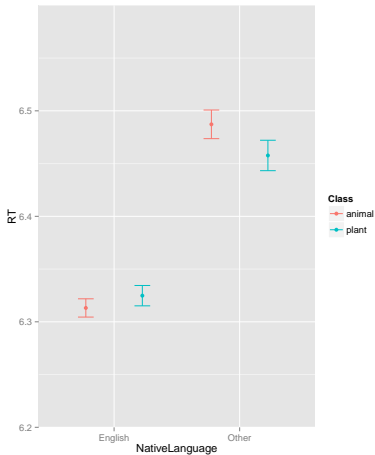
LI539
Mixed
Effect
Models
Degen

```
ggplot(agr, aes(x=NativeLanguage, y=RT, fill=Class)) +  
  geom_bar(stat="identity", position=dodge) +  
  geom_errorbar(aes(ymin=YMin, ymax=YMax), width=0.25, position=dodge) +  
  coord_cartesian(ylim=c(6, 6.6))
```



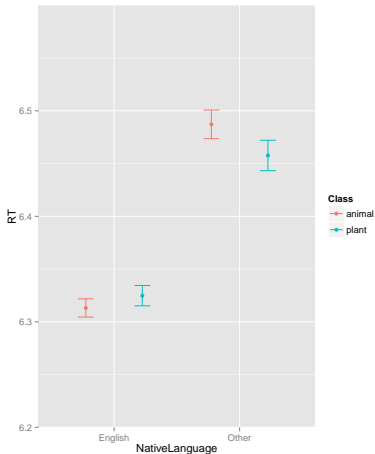
Error bars on points.

- create the same plot using dots instead of bars



Error bars on points.

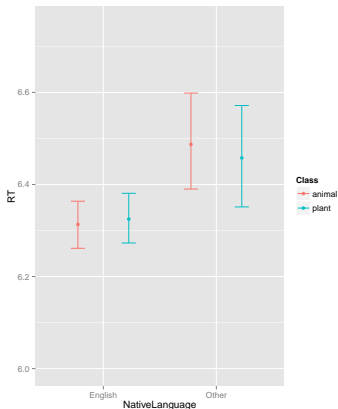
- create the same plot using dots instead of bars



```
ggplot(agr, aes(x=NativeLanguage, y=RT, color=Class)) +  
  geom_point(position=dodge) +  
  geom_errorbar(aes(ymin=YMin, ymax=YMax), width=0.25, position=dodge) +  
  coord_cartesian(ylim=c(6.2, 6.6))
```


Error bars on points

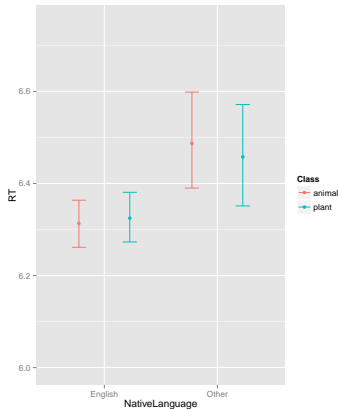
- create the same plot but with error bars representing 95% confidence intervals



```
agrr = aggregate(RT ~ NativeLanguage +  
  Class + Subject, lexdec, mean)  
agr = aggregate(RT ~ NativeLanguage +  
  Class, agrr, mean)  
agr$CIH = aggregate(RT ~ NativeLanguage +  
  Class, agrr, ci.high)$RT  
agr$CIL = aggregate(RT ~ NativeLanguage +  
  Class, agrr, ci.low)$RT  
agr$YMin = agr$RT - agr$CIL  
agr$YMax = agr$RT + agr$CIH
```

Error bars on points

- create the same plot but with error bars representing 95% confidence intervals



```
agrr = aggregate(RT ~ NativeLanguage +  
  Class + Subject, lexdec, mean)  
agr = aggregate(RT ~ NativeLanguage +  
  Class, agrr, mean)  
agr$CIH = aggregate(RT ~ NativeLanguage +  
  Class, agrr, ci.high)$RT  
agr$CIL = aggregate(RT ~ NativeLanguage +  
  Class, agrr, ci.low)$RT  
agr$YMin = agr$RT - agr$CIL  
agr$YMax = agr$RT + agr$CIH
```

```
ggplot(agr, aes(x=NativeLanguage, y=RT, color=Class)) +  
  geom_point(position=dodge) +  
  geom_errorbar(aes(ymin=YMin, ymax=YMax), width=0.25,  
    position=dodge) +  
  scale_y_continuous(limits=c(6, 6.75))
```

Comparison of error bars

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

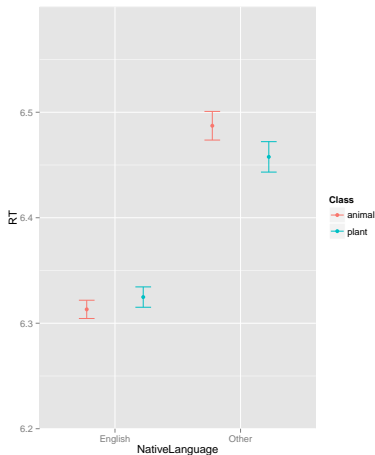
Histograms
and
densities

Bar plots

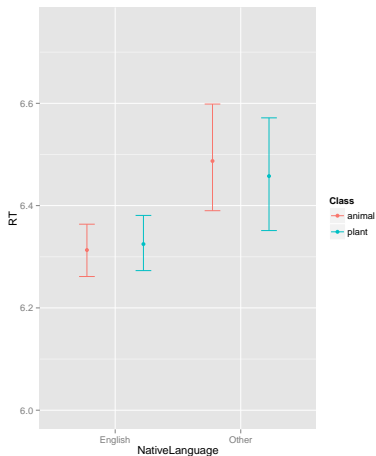
Scatterplots

Manipulating
plot
appearance

• 1 SE



• 95% confidence intervals

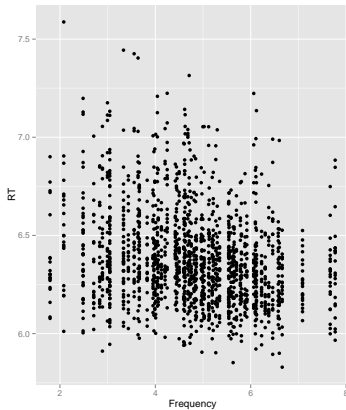


Scatterplots

LI539
Mixed
Effect
Models

Degen

- create scatterplot of RTs by frequency



What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

Scatterplots

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

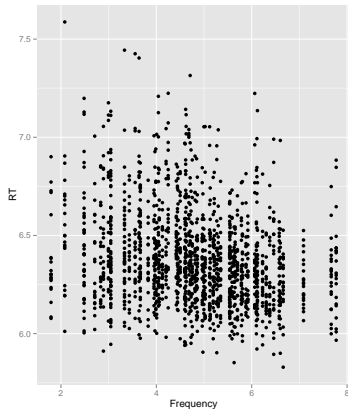
Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- create scatterplot of RTs by frequency



```
ggplot(lexdec, aes(x=Frequency, y=RT)) +  
  geom_point()
```

Jittered scatterplots

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

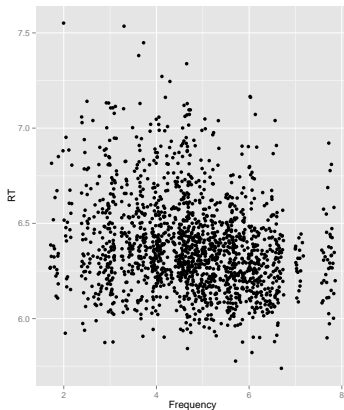
Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- jitter the points to increase readability



Jittered scatterplots

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

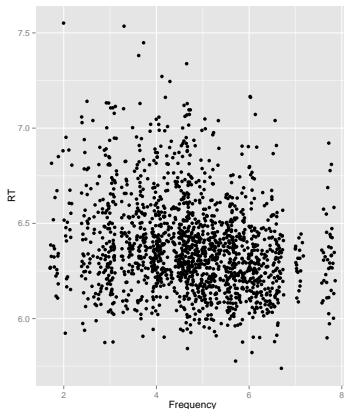
Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

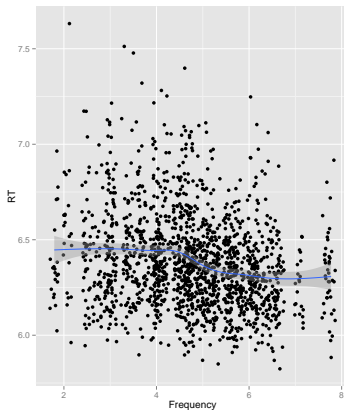
- jitter the points to increase readability



```
ggplot(lexdec, aes(x=Frequency, y=RT)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1))
```

Smoothers I

- fit a lowess line through the data cloud



Smoothers I

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

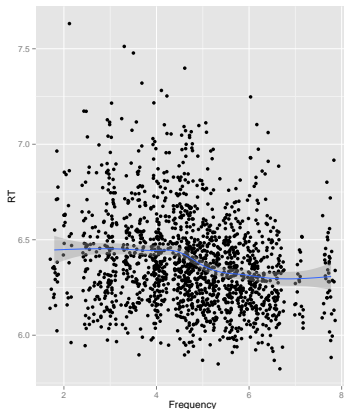
Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

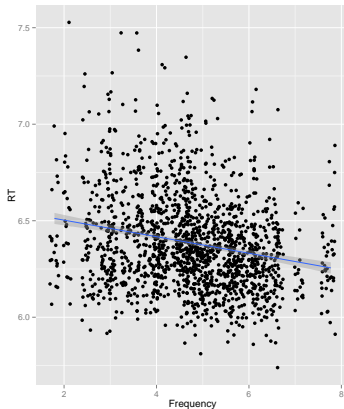
- fit a lowess line through the data cloud



```
ggplot(lexdec, aes(x=Frequency, y=RT)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1)) +  
  stat_smooth()
```

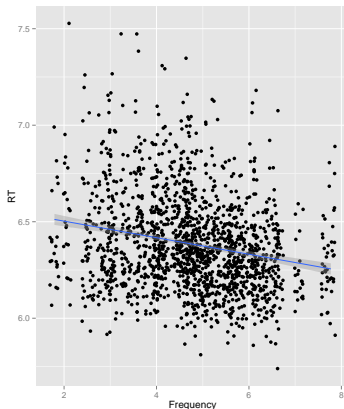
Smoothers II

- fit a linear smoother through the data cloud



Smoothers II

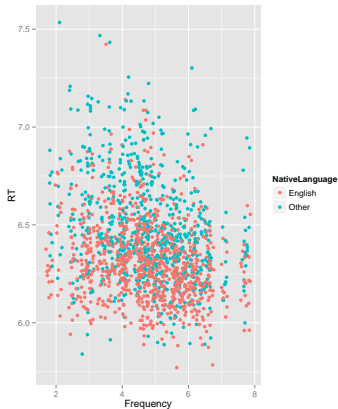
- fit a linear smoother through the data cloud



```
ggplot(lexdec, aes(x=Frequency, y=RT)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1)) +  
  stat_smooth(method="lm")
```

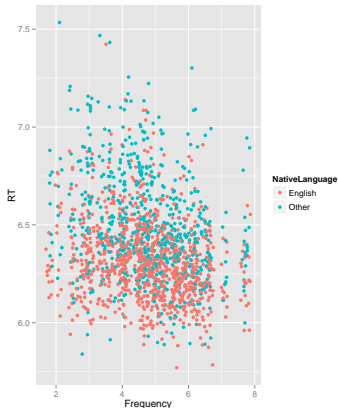
Color in scatterplots

- create the same scatterplot with colors representing different native languages



Color in scatterplots

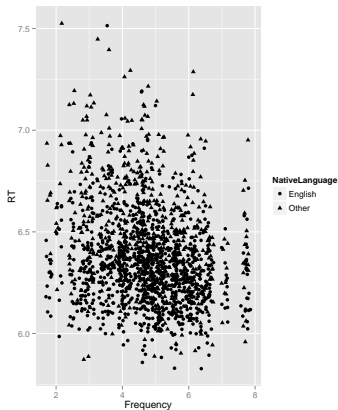
- create the same scatterplot with colors representing different native languages



```
ggplot(lexdec, aes(x=Frequency, y=RT, color=NativeLanguage)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1))
```

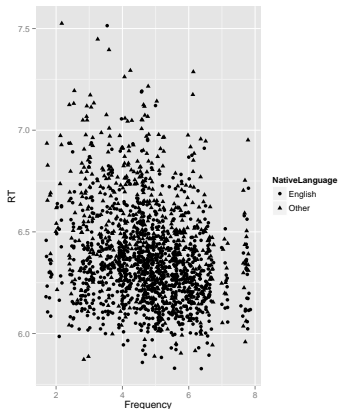
Shapes in scatterplots

- create the same scatterplot with shapes representing different native languages



Shapes in scatterplots

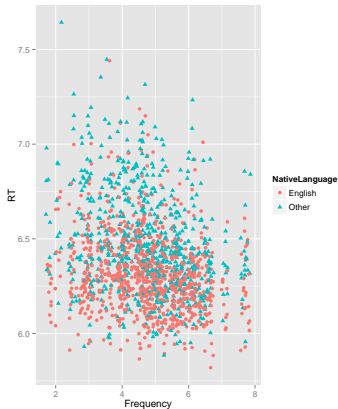
- create the same scatterplot with shapes representing different native languages



```
ggplot(lexdec, aes(x=Frequency, y=RT, shape=NativeLanguage)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1))
```

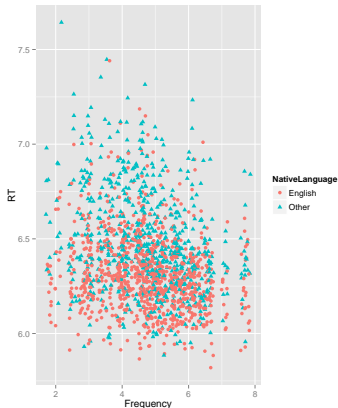
Multiple aesthetics in scatterplots

- create the same scatterplot with shapes **and** color representing different native languages



Multiple aesthetics in scatterplots

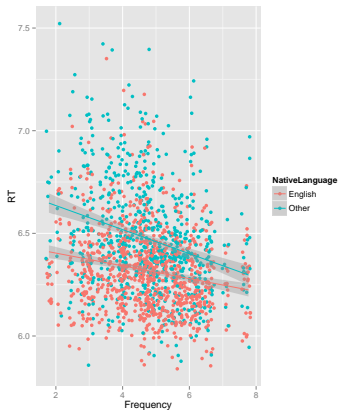
- create the same scatterplot with shapes **and** color representing different native languages



```
ggplot(lexdec, aes(x=Frequency, y=RT, shape=NativeLanguage, color=NativeLanguage)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1))
```

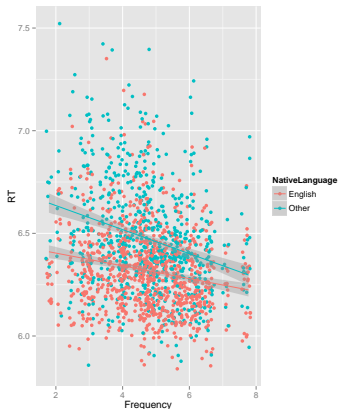
Multiple smoothers in scatterplots

- add separate linear smoothers for different native languages



Multiple smoothers in scatterplots

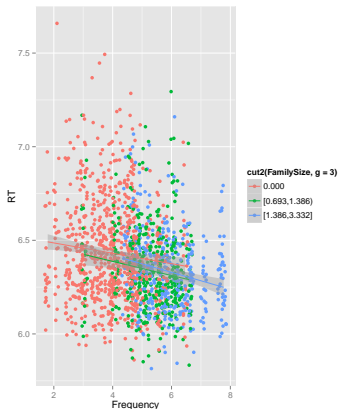
- add separate linear smoothers for different native languages



```
ggplot(lexdec, aes(x=Frequency, y=RT, color=NativeLanguage)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1)) +  
  stat_smooth(method="lm")
```

Three-way continuous interactions

- turn one of the continuous variables into a factor by cutting it into bins of equal sizes (i.e. containing as close to as possible equal number of data points) and
- example: cut the FamilySize variables into 3 bins
- then plot binned FamilySize on top of the scatterplot



Binning a variable

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- use `cut` (equally sized intervals) or `cut2` (bins of roughly equal cardinality)

```
table(cut(lexdec$FamilySize,breaks=3))
```

```
##
```

```
## (-0.00333,1.11]      (1.11,2.22]      (2.22,3.34]
```

```
##           1260           294           105
```

```
table(cut2(lexdec$FamilySize,g=3))
```

```
##
```

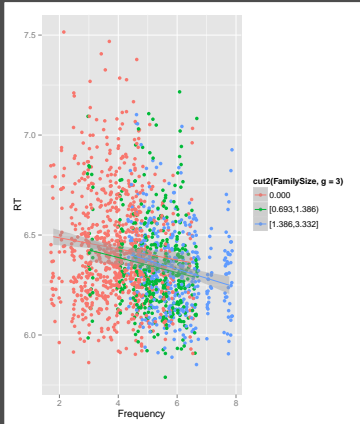
```
##      0.000 [0.693,1.386) [1.386,3.332]
```

```
##      840           420           399
```

Three-way continuous interactions

- include the cut2 call directly in the ggplot call

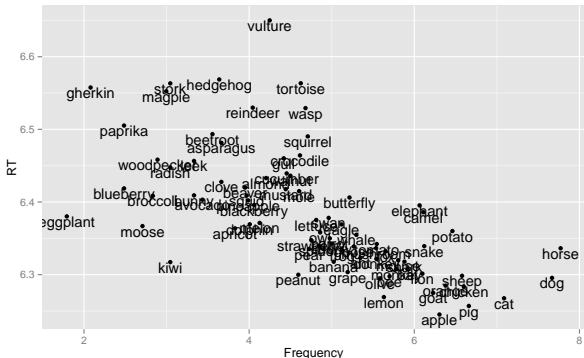
```
ggplot(lexdec, aes(x=Frequency, y=RT, color=cut2(FamilySize,g=3))) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1)) +  
  stat_smooth(method="lm")
```



Annotating a scatterplot

- compute mean RTs for each word
- create scatterplot of mean RTs by frequency
- label each point with the word it represents

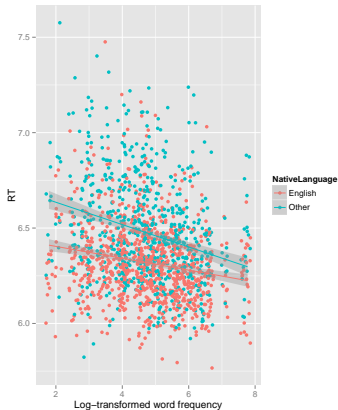
```
agr = aggregate(RT ~ Word + Frequency, lexdec, mean)
ggplot(agr, aes(x=Frequency, y=RT)) +
  geom_point() +
  geom_text(aes(label=Word), vjust=1)
```



Manipulating plot appearance: scales

- scales control aspects of axes, colors, shapes, etc.
- e.g. relabel x-axis

```
ggplot(lexdec, aes(x=Frequency, y=RT, color=NativeLanguage)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1)) +  
  stat_smooth(method="lm") +  
  scale_x_continuous(name="Log-transformed word frequency")
```

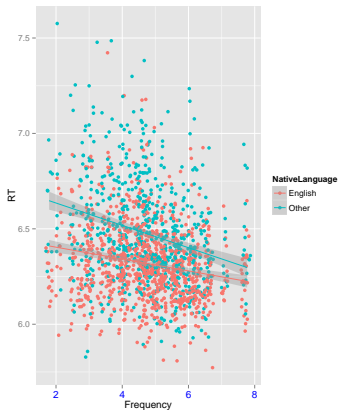


Relabel other axis

The theme() function

- control label sizes/colors; plot background; legend features
- e.g., increase the size of the x axis labels and change their color

```
ggplot(lexdec, aes(x=Frequency, y=RT, color=NativeLanguage)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1)) +  
  stat_smooth(method="lm") +  
  theme(axis.text.x=element_text(size=12,color="blue"))
```



The theme() function

LI539
Mixed
Effect
Models
Degen

What is
ggplot?

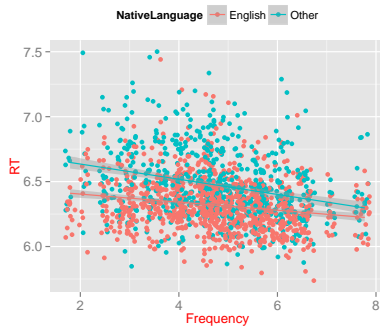
Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- increase the size of both x and y-axis labels simultaneously
- change the color of axis titles
- move the legend to the top of the plot.



The theme() function

L1539
Mixed
Effect
Models
Degen

What is
ggplot?

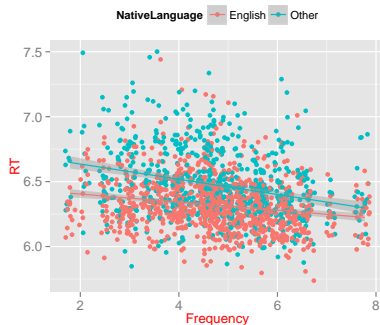
Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

- increase the size of both x and y-axis labels simultaneously
- change the color of axis titles
- move the legend to the top of the plot.



```
ggplot(lexdec, aes(x=Frequency, y=RT, color=NativeLanguage)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1)) +  
  stat_smooth(method="lm") +  
  theme(axis.text=element_text(size=12), axis.title=  
    element_text(color="red"), legend.position="top")
```

Background (white good for publications)

LI539
Mixed
Effect
Models

Degen

What is
ggplot?

Histograms
and
densities

Bar plots

Scatterplots

Manipulating
plot
appearance

```
theme_set(theme_bw())  
ggplot(lexdec, aes(x=Frequency, y=RT, color=NativeLanguage)) +  
  geom_jitter(position = position_jitter(width = 0.1, height=0.1)) +  
  stat_smooth(method="lm") +  
  theme(axis.text=element_text(size=12), axis.title=  
    element_text(color="red"))
```

