

Making graphs with ggplot2

Hugh Murrell

University of KwaZulu-Natal
Computer Science
Pietermaritzburg

Honours, Data Mining
February - May, 2016

based on slides by Brian Diggs.

Outline

1 Background and Philosophy

- Statistical Graphics
- Philosophical Structure

2 Elements of the Grammar

- Initializer
- Geoms
- Layers
- Facets
- Groups
- Stats
- Exercise

Designed for statistical or data graphics: where quantities of the data are represented by position, shape, color, etc.

Not designed for

- Organization charts
- Flow charts
- Infographics
- Posters

Philosophical Structure

- Data in a data.frame in “long” format
- Columns of data are mapped to aesthetics
- Non-data related aspects controlled by theme

Data in long format

data.frame with an entry for each measurement and the variables associated with describing the circumstances of that measurement

```
> names(ChickWeight)

[1] "weight" "Time"    "Chick"
[4] "Diet"
```

Aesthetics

- Examples
 - Position along an axis
 - Shape
 - Color
 - Width
 - Line type
 - Replicate (horizontal and/or vertical)
- Legend is mapping between visual element and the data values
- Any aesthetic can be fixed to a specific (non-data) value

Themes

Themes control non-data related aspects of the graphic

- Size of title
- Font of tick labels
- Location of legend
- Background colors

qplot vs. ggplot

Two ways to start making a plot

- qplot
 - Designed to be like plot
 - Eases transition
 - Obscures details
- ggplot
 - Core of the actual grammar
 - Less familiar

ggplot

Arguments

data Default data.frame of the data to be plotted

mapping Default aesthetic mappings

```
> ggplot(ChickWeight)
```

Error: No layers in plot

Incomplete by itself

Geoms

Geoms are the different ways that “ink” is used to show the values

- Points
- Lines connecting points/locations
- Steps connecting points/locations
- Bars
- Points with lines extending from them (error bars)
- Tiles/rectangles
- Ribbons
- Text

Required and optional aesthetics

Geom	Required	Optional
point	x, y (positions)	shape, colour, size, fill, alpha
rect	xmin, xmax, ymin, ymax	colour, fill, size, line-type, alpha
text	x, y, label	colour, size, angle, hjust, vjust, alpha

Aesthetics are specified by calls to the `aes` function with arguments that are aesthetic name and the variable being mapped to it.

Examples of geoms

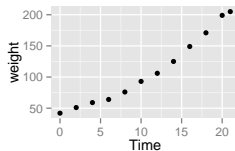
`cw` is a subset of the `ChickWeight` data.frame in `datasets`

```
> library(dplyr)
> cw <- ChickWeight %>% filter(Chick==1)

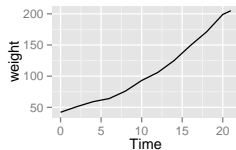
> ggplot(cw) + geom_point(aes(x=Time, y=weight))
> ggplot(cw) + geom_line(aes(x=Time, y=weight))
> ggplot(cw) + geom_path(aes(x=Time, y=weight))
> ggplot(cw) + geom_step(aes(x=Time, y=weight))
```

Examples of geoms (results)

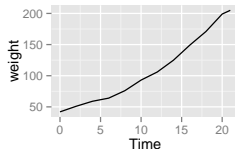
geom_point



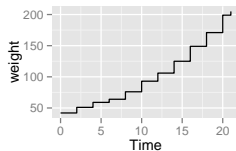
geom_line



geom_path



geom_step

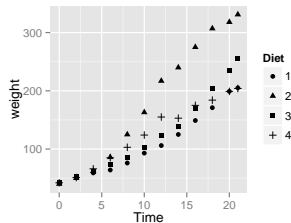
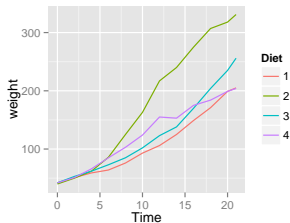


Optional aesthetics

`cw2` is a broader subset of `ChickWeight` containing data from the first chick on each of the four possible diets.

```
> cw2 <- ChickWeight %>% group_by(Diet) %>%  
+   filter(Chick==Chick[1])  
  
> ggplot(cw2) +  
+   geom_line(aes(x=Time, y=weight, colour=Diet))  
  
> ggplot(cw2) +  
+   geom_point(aes(x=Time, y=weight, shape=Diet))
```

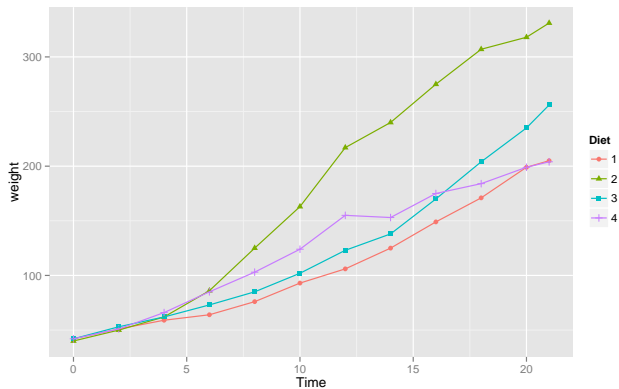
Optional aesthetics (results)



Layers

Much of the flexibility comes from being able to combine different geoms in layers

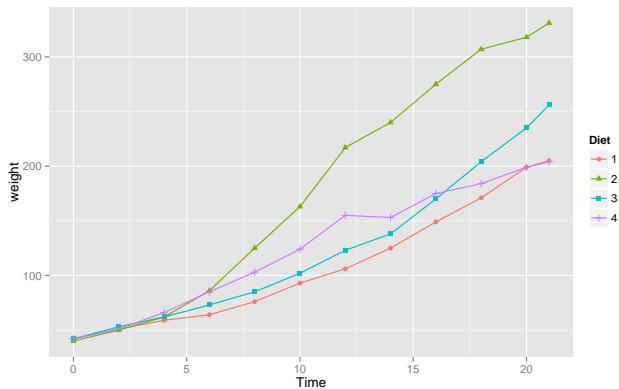
```
> ggplot(cw2) +  
+   geom_point(aes(x=Time, y=weight,  
+                 colour=Diet, shape=Diet)) +  
+   geom_line(aes(x=Time, y=weight,  
+                 colour=Diet))
```

Default aesthetics

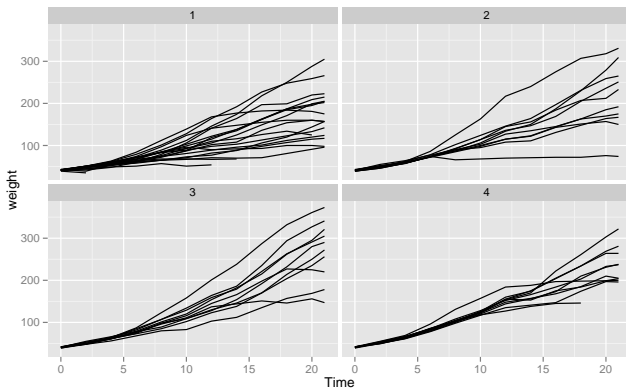
Multiple geoms using the same set of aesthetics show the utility of default aesthetics

```
> ggplot(cw2, aes(x=Time, y=weight, colour=Diet)) +  
+   geom_point(aes(shape=Diet)) +  
+   geom_line()
```



Facets, or small multiples, are sets of graphs showing subsets of data on common x and/or y axes for comparison purposes

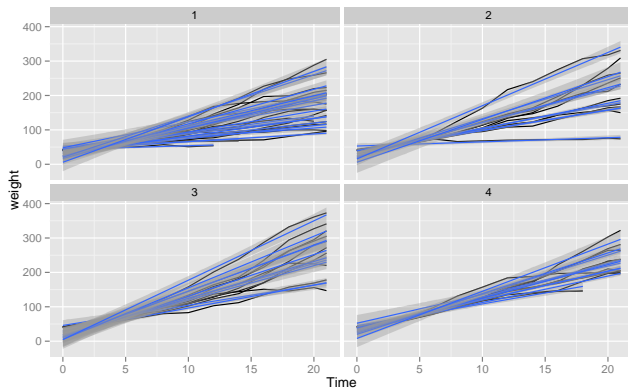
```
> ggplot(ChickWeight,  
+       aes(x=Time, y=weight, group=Chick)) +  
+       geom_line() +  
+       facet_wrap(~Diet)  
>
```



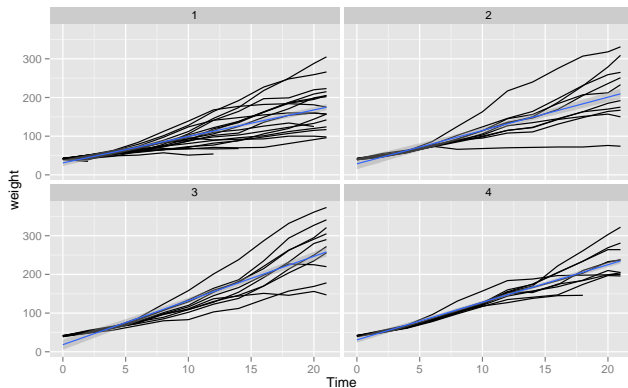
- By default, a separate group is created for each combination of categorical (factor or character) variables that are mapped to aesthetics
- Can be overridden with the group aesthetic
- Many things done on a one-per-group basis (line, path, ribbon, stat)

- Transformation (summarization) of data by group
 - Binning (1 or 2 dimensional)
 - Quartiles
 - Mean and standard deviation
 - Smoothing/regression
- Results of the transformations are mapped to aesthetics

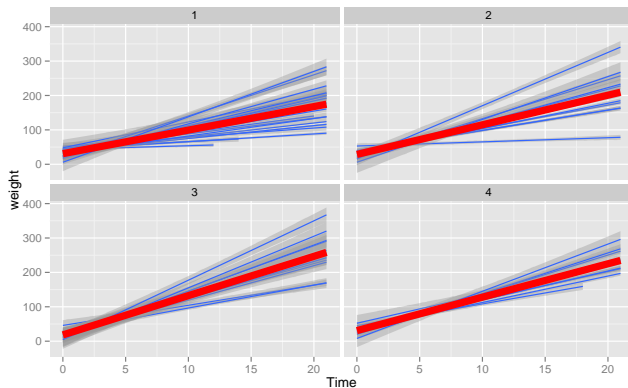
```
> ggplot(ChickWeight,  
+       aes(x=Time, y=weight, group=Chick)) +  
+       geom_line() +  
+       stat_smooth(method="lm") +  
+       facet_wrap(~Diet)  
>
```

```
> ggplot(ChickWeight,  
+       aes(x=Time, y=weight, group=Chick)) +  
+       geom_line() +  
+       stat_smooth(aes(group=Diet), method="lm") +  
+       facet_wrap(~Diet)  
>
```



```
> ggplot(ChickWeight,  
+       aes(x=Time, y=weight, group=Chick)) +  
+       stat_smooth(method="lm") +  
+       stat_smooth(aes(group=Diet),  
+                   method="lm",  
+                   colour="red", size=3) +  
+       facet_wrap(~Diet)  
>
```



Possible question for second test.

Make use of the *ggplot2* package to re-create the tree plots from the weeks 3 and 4 slides.

Try and make use of *facets* to create growth plots for all trees in a forest on a forest by forest basis.

Annotate your forest growth plots with the average growth rate per forest.