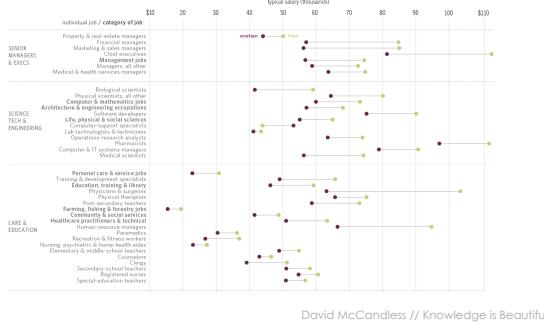
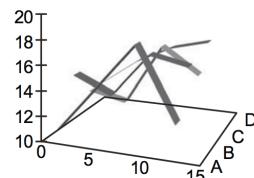


Tell a story with your data

Gender Pay Gap US



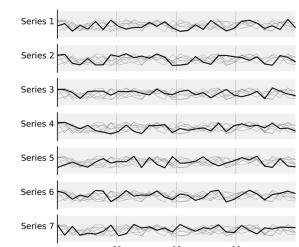
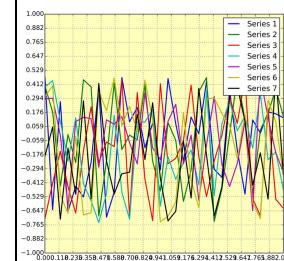
David McCandless // Knowledge is Beautiful



"Create the simplest graph that conveys the information that you want to convey."

Edward Tufte // The Visual Display of Quantitative Information

Eliminating "chartjunk"

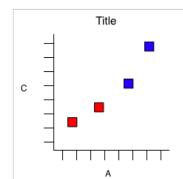


What's in a plot?

The layered grammar of graphics.

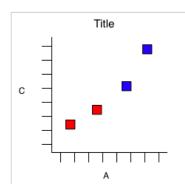
Grammar of graphics

- What is a plot?
- How can we succinctly describe a plot?
- How can we create a plot that we have described?



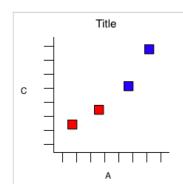
How do we build a plot?

- Start with data
 - A = predictor variable
 - B = modifier variable (e.g. treatment)
 - C = response variable



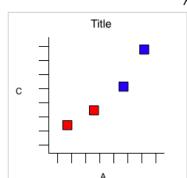
What are the components of a plot?

- Data is represented by geometric objects
- Geometric objects are positioned using scales and coordinate systems
- Meaning is given through plot annotations



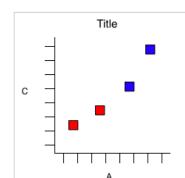
Plot layers

Geometric objects Scales & coordinates Annotations



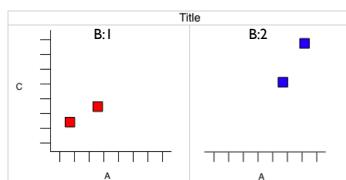
Plot layers

- How else could we display this visually?



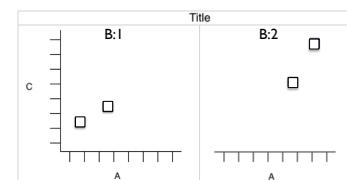
Plot layers

- How else could we display this visually?



Plot layers

- How else could we display this visually?
 - remember to remove 'chartjunk'...



{ggplot2}

Using the layered grammar of graphics.

Building a plot with {ggplot2}

- Scatterplot of biomass against slope from worms data

FIELD.NAME	BIO MASS	SLOPE	VEGETATION	SOIL.PH	DAMP	WORM.DENSITY
1 Nashs.Field	3.6	11	Grassland	4.1	FALSE	4
2 Silwood.Bottom	5.1	2	Arable	5.2	FALSE	7
3 Nursery.Field	2.8	3	Grassland	4.3	FALSE	2
4 Rush.Meadow	2.4	5	Meadow	4.9	TRUE	5
5 Gunness.Thicket	3.8	0	Scrub	4.2	FALSE	6
6 Oak.Mead	3.1	2	Grassland	3.9	FALSE	2
7 Church.Field	3.5	3	Grassland	4.2	FALSE	3
8 Ashurst	2.1	0	Arable	4.8	FALSE	4
9 The.Orchard	1.9	0	Orchard	5.7	FALSE	9
10 Rookery.Slope	1.5	4	Grassland	5.0	TRUE	7

Building a plot with {ggplot2}

- Scatterplot of biomass against slope from worms data
 - Data: worms

FIELD.NAME	BIO MASS	SLOPE	VEGETATION	SOIL.PH	DAMP	WORM.DENSITY
1 Nashs.Field	3.6	11	Grassland	4.1	FALSE	4
2 Silwood.Bottom	5.1	2	Arable	5.2	FALSE	7
3 Nursery.Field	2.8	3	Grassland	4.3	FALSE	2
4 Rush.Meadow	2.4	5	Meadow	4.9	TRUE	5
5 Gunness.Thicket	3.8	0	Scrub	4.2	FALSE	6
6 Oak.Mead	3.1	2	Grassland	3.9	FALSE	2
7 Church.Field	3.5	3	Grassland	4.2	FALSE	3
8 Ashurst	2.1	0	Arable	4.8	FALSE	4
9 The.Orchard	1.9	0	Orchard	5.7	FALSE	9
10 Rookery.Slope	1.5	4	Grassland	5.0	TRUE	7

Building a plot with {ggplot2}

- Scatterplot of biomass against slope from worms data
 - Data: worms
 - X: slope

FIELD.NAME	BIO MASS	SLOPE	VEGETATION	SOIL.PH	DAMP	WORM.DENSITY
1 Nashs.Field	3.6	11	Grassland	4.1	FALSE	4
2 Silwood.Bottom	5.1	2	Arable	5.2	FALSE	7
3 Nursery.Field	2.8	3	Grassland	4.3	FALSE	2
4 Rush.Meadow	2.4	5	Meadow	4.9	TRUE	5
5 Gunness.Thicket	3.8	0	Scrub	4.2	FALSE	6
6 Oak.Mead	3.1	2	Grassland	3.9	FALSE	2
7 Church.Field	3.5	3	Grassland	4.2	FALSE	3
8 Ashurst	2.1	0	Arable	4.8	FALSE	4
9 The.Orchard	1.9	0	Orchard	5.7	FALSE	9
10 Rookery.Slope	1.5	4	Grassland	5.0	TRUE	7

Building a plot with {ggplot2}

- Scatterplot of biomass against slope from worms data
 - Data: worms
 - X: slope
 - Y: biomass

FIELD.NAME	BIO MASS	SLOPE	VEGETATION	SOIL.PH	DAMP	WORM.DENSITY
1 Nashs.Field	3.6	11	Grassland	4.1	FALSE	4
2 Silwood.Bottom	5.1	2	Arable	5.2	FALSE	7
3 Nursery.Field	2.8	3	Grassland	4.3	FALSE	2
4 Rush.Meadow	2.4	5	Meadow	4.9	TRUE	5
5 Gunness.Thicket	3.8	8	Scrub	4.2	FALSE	6
6 Oak.Head	3.1	2	Grassland	3.9	FALSE	2
7 Church.Field	3.5	3	Grassland	4.2	FALSE	3
8 Ashurst	2.1	8	Arable	4.8	FALSE	4
9 The.Orchard	1.9	8	Orchard	5.7	FALSE	9
10 Rookery.Slope	1.5	4	Grassland	5.8	TRUE	7

Layered grammar with {ggplot2}

- Scatterplot of biomass against slope from worms data
 - Specify default **data** and variables to be mapped into **aesthetic** space

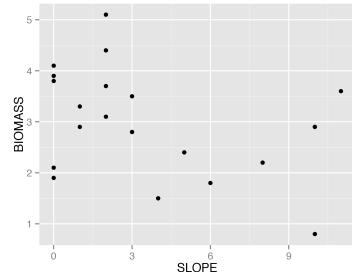
```
ggplot(data = WORMS,
       aes(x = SLOPE,
            y = BIOMASS))
```

Layered grammar with {ggplot2}

- Scatterplot of biomass against slope from worms data
 - Specify default **data** and variables to be mapped into **aesthetic** space
 - Specify how our **geometric** objects are to be shown

```
ggplot(data = WORMS,
       aes(x = SLOPE,
            y = BIOMASS)) +
       geom_point()
```

Layered grammar with {ggplot2}



Building a plot with {ggplot2}

- Scatterplot of biomass against slope from worms data
 - Data: worms
 - X: slope
 - Y: biomass
 - Colour: damp

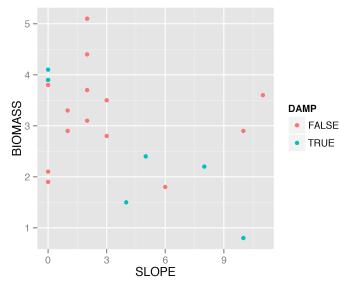
FIELD.NAME	BIO MASS	SLOPE	VEGETATION	SOIL PH	DAMP	WORM.DENSITY
1 Nashes.Field	3.6	11	Grassland	4.1	FALSE	4
2 Silwood.Bottom	5.1	2	Arable	5.2	FALSE	7
3 Nursery.Field	2.8	3	Grassland	4.3	FALSE	2
4 Rush.Meadow	2.4	5	Meadow	4.9	TRUE	5
5 Gunness.Thicket	3.8	8	Scrub	4.2	FALSE	6
6 Oak.Head	3.1	2	Grassland	3.9	FALSE	2
7 Church.Field	3.5	3	Grassland	4.2	FALSE	3
8 Ashurst	2.1	6	Arable	4.8	FALSE	4
9 The.Orchard	1.9	8	Orchard	5.7	FALSE	9
10 Rookery.Slope	1.5	4	Grassland	5.8	TRUE	7

Layered grammar with {ggplot2}

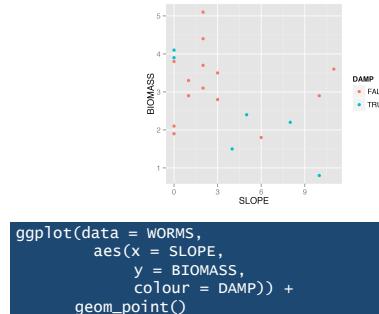
- Scatterplot of biomass against slope from worms data, conditional on damp
 - Specify default **data** and variables to be mapped into **aesthetic** space
 - Specify how our **geometric** objects are to be shown

```
ggplot(data = WORMS,
       aes(x = SLOPE,
            y = BIOMASS,
            colour = DAMP)) +
  geom_point()
```

Layered grammar with {ggplot2}



Layered grammar with {ggplot2}



```
ggplot(data = WORMS,
       aes(x = SLOPE,
            y = BIOMASS,
            colour = DAMP)) +
  geom_point()
```

geoms

Exploring geometric object types.

Histogram

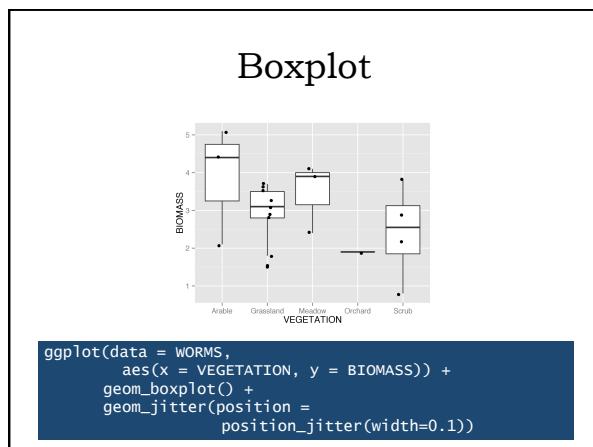
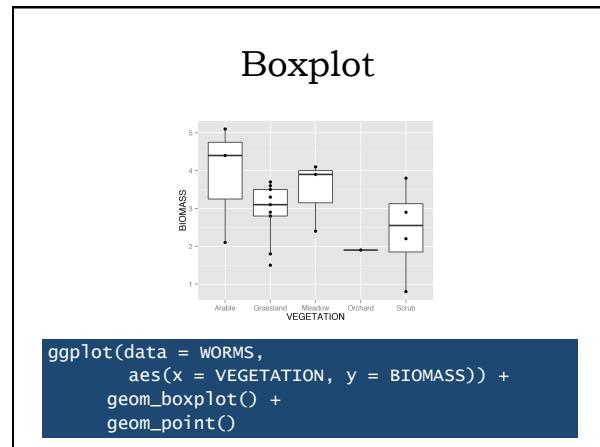
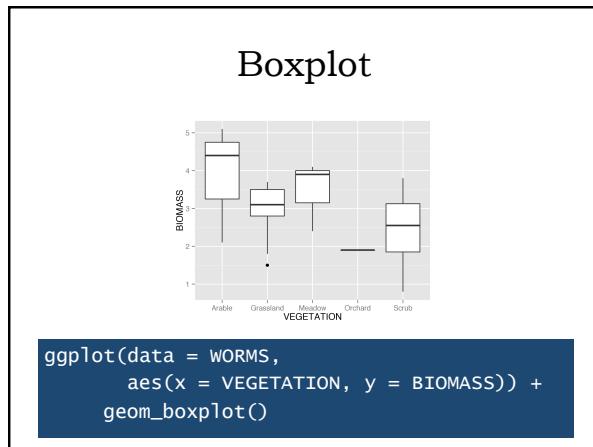
```
ggplot(data = WORMS, aes(x = BIOMASS)) +
  geom_histogram()
```

Histogram

```
ggplot(data = WORMS, aes(x = BIOMASS)) +
  geom_histogram(binwidth = 0.7)
```

Histogram

```
ggplot(data = WORMS, aes(x = BIOMASS)) +
  geom_histogram(binwidth = 0.7,
                 fill = 'white', colour = 'black')
```

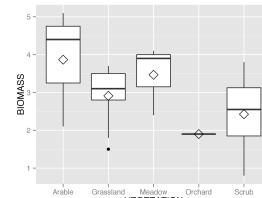


stats
Statistical transformations.

Statistical transformations

- Transforming your data:
 - bin: divide continuous range into bins and count
 - summary: aggregate values of y for given x
 - smooth: smoothed conditional mean of y given x

Boxplot with group means

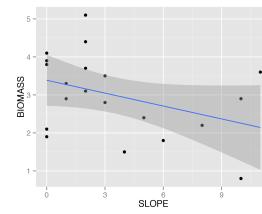


```
ggplot(data = WORMS,
       aes(x = VEGETATION, y = BIOMASS)) +
  geom_boxplot() +
  stat_summary(fun.y = "mean",
              geom = "point",
              shape = 5, size = 4)
```

Statistical transformations

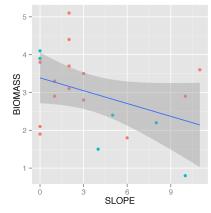
- Transforming your data:
 - bin: divide continuous range into bins and count
 - summary: aggregate values of y for given x
 - smooth: smoothed conditional mean of y given x

Scatterplot with linear regression



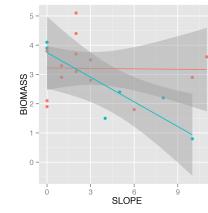
```
ggplot(data = WORMS,
       aes(x = VEGETATION, y = BIOMASS)) +
  geom_point() +
  stat_smooth(method = "lm")
```

Scatterplot with linear regression



```
ggplot(data = WORMS,
       aes(x = VEGETATION, y = BIOMASS)) +
  geom_point(aes(colour = DAMP)) +
  stat_smooth(method = "lm")
```

Scatterplot with linear regression



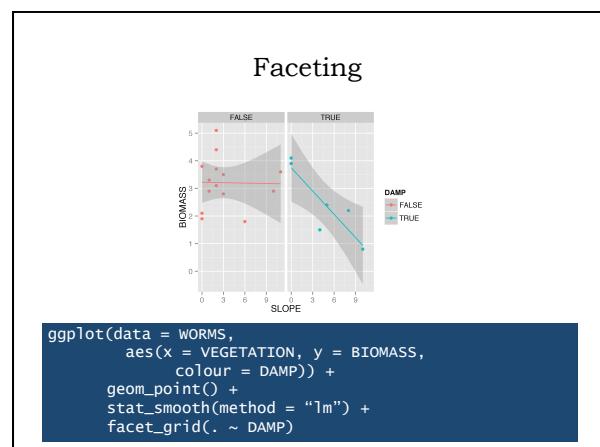
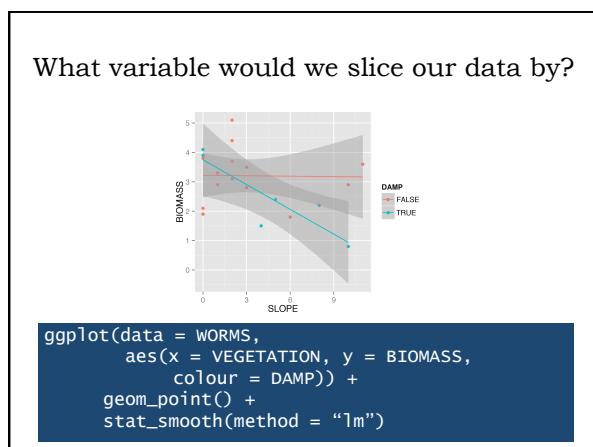
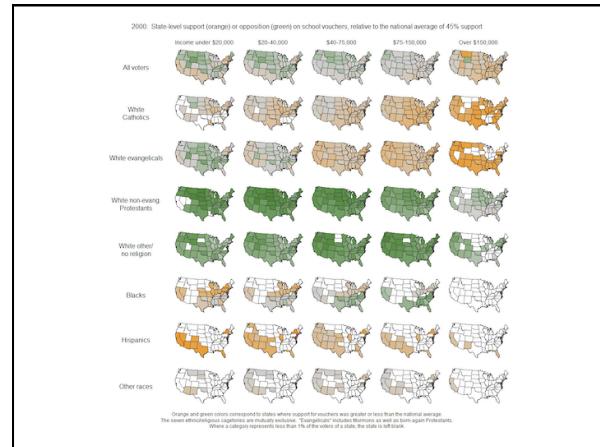
```
ggplot(data = WORMS,
       aes(x = VEGETATION, y = BIOMASS,
           colour = DAMP)) +
  geom_point() +
  stat_smooth(method = "lm")
```

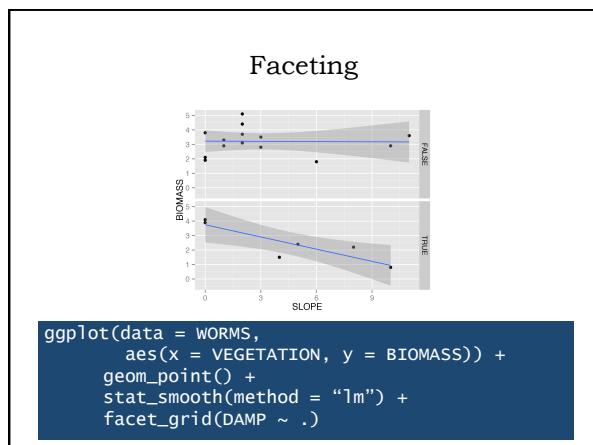
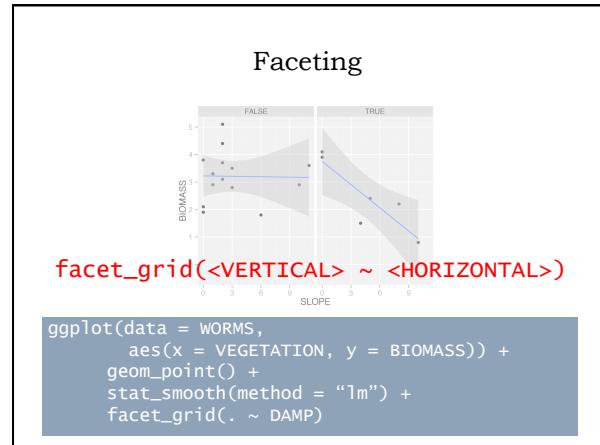
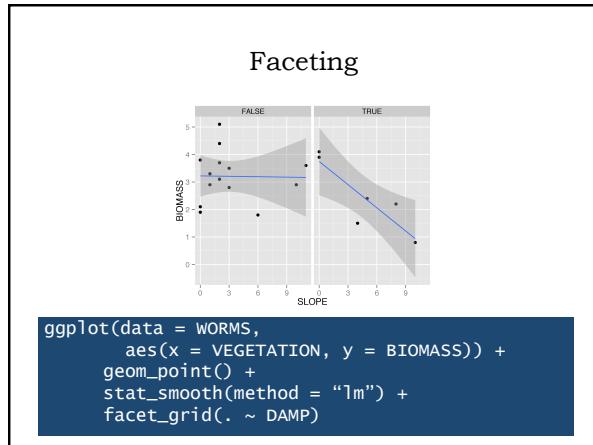
Faceting

Small multiples.

Small multiples

- Use the same basic graphic or chart to display different slices of a data set.
- Present multi-dimensional data without cramming into an overly-complex figure.





Practical exercise.

Plotting data with {ggplot2}.



Useful links

- ggplot2 documentation
 - <http://docs.ggplot2.org/current/>
- R graphics cookbook
 - <http://www.cookbook-r.com/Graphs/>