Recap Location Tests

Mandy Vogel

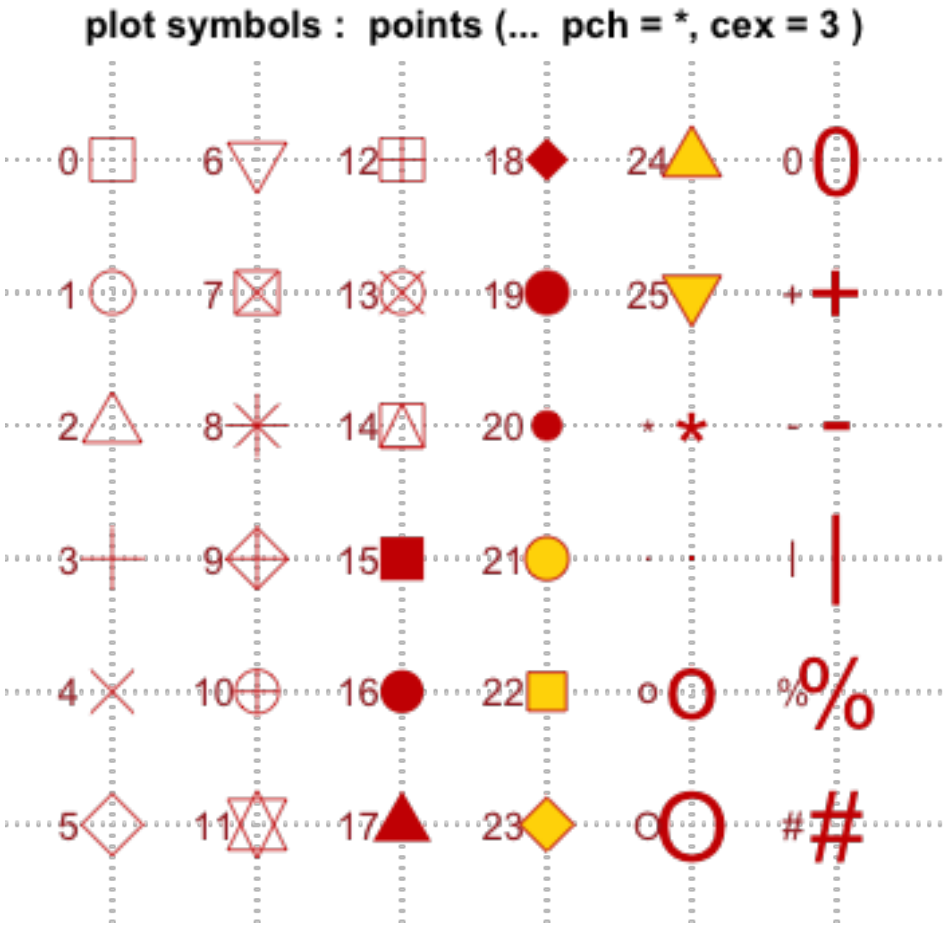
July 6, 2015

Table of Contents

Scatter plots I

- up to now we only visualized univariate numerical data but we already used a central geom for visualizing bivariate data
- to produce a scatter plot we use `geom_point()` with the aesthetics `aes(x = variable1, y = variable2)`
- we can customize the size (`size`), shape (`shape`), colour(`colour` and `fill`), and transparency (`alpha`) of the points

Scatter plots II

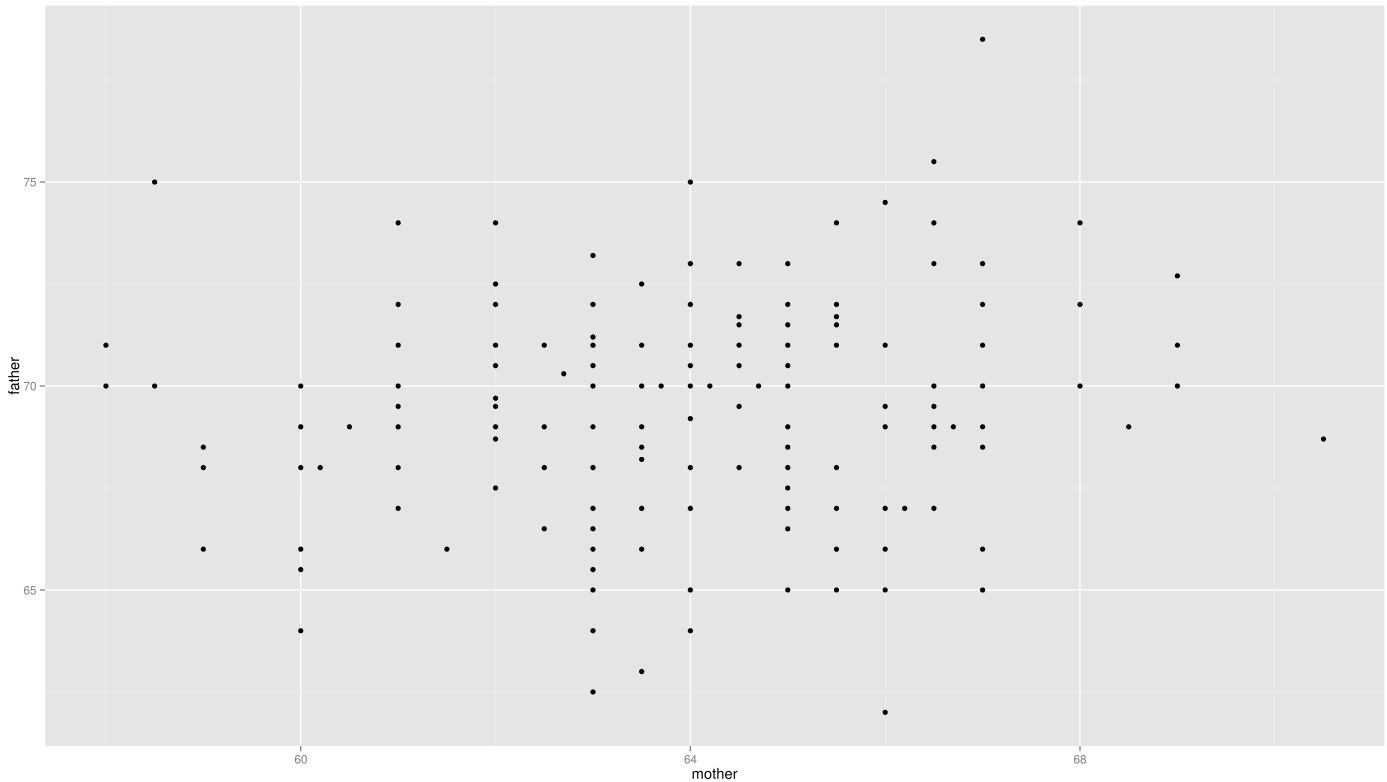


Scatter plots

```
> ggplot(GaltonFamilies, aes(x=mother,y=father)) +
+     geom_point()
```

which results in:

Scatter plots



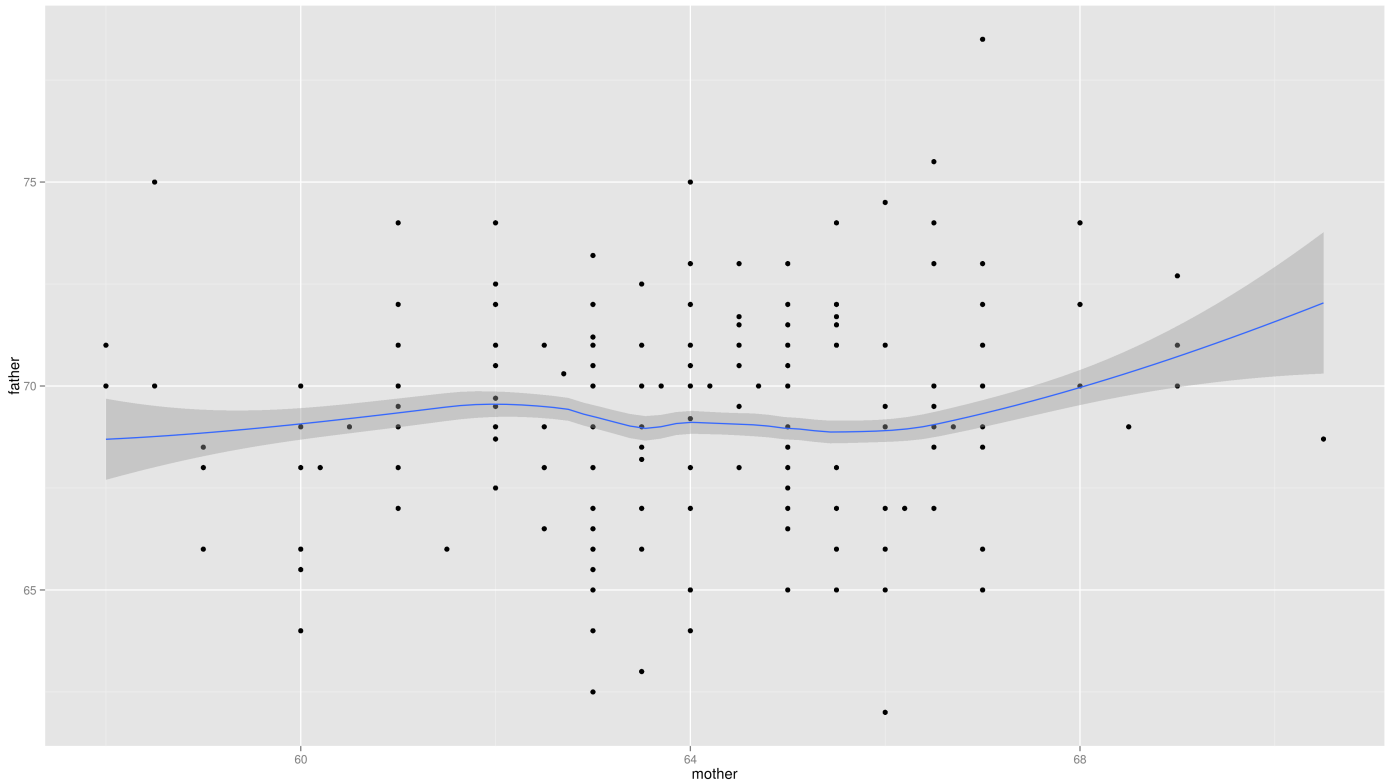
Adding trend lines

- a trend line is a statistical summary of a bivariate relationship
- there are many different trend lines that can be added to a scatterplot
- they are easily added through `geom_smooth()`

```
> ggplot(GaltonFamilies, aes(x=mother,y=father)) +
+     geom_point()
```

which results in:

Adding trend lines



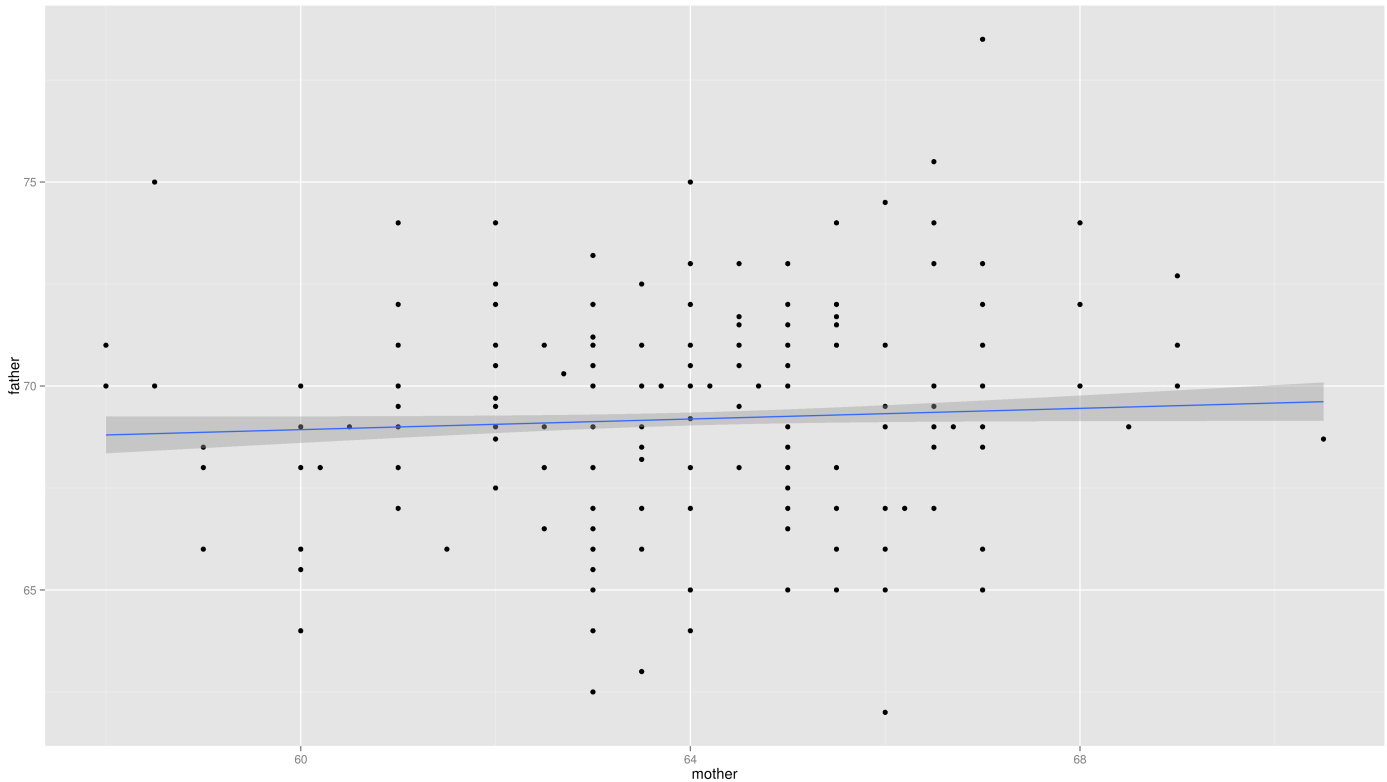
Adding trend lines

- by default `geom_smooth()` produces a non-parametric trend line (local polynomial regression fitting for n¡1000 otherwise generalized additive model)
- we can set a model using the `method` argument

```
> ggplot(GaltonFamilies, aes(x=mother,y=father)) +
+     geom_point()
```

which results in:

Adding trend lines



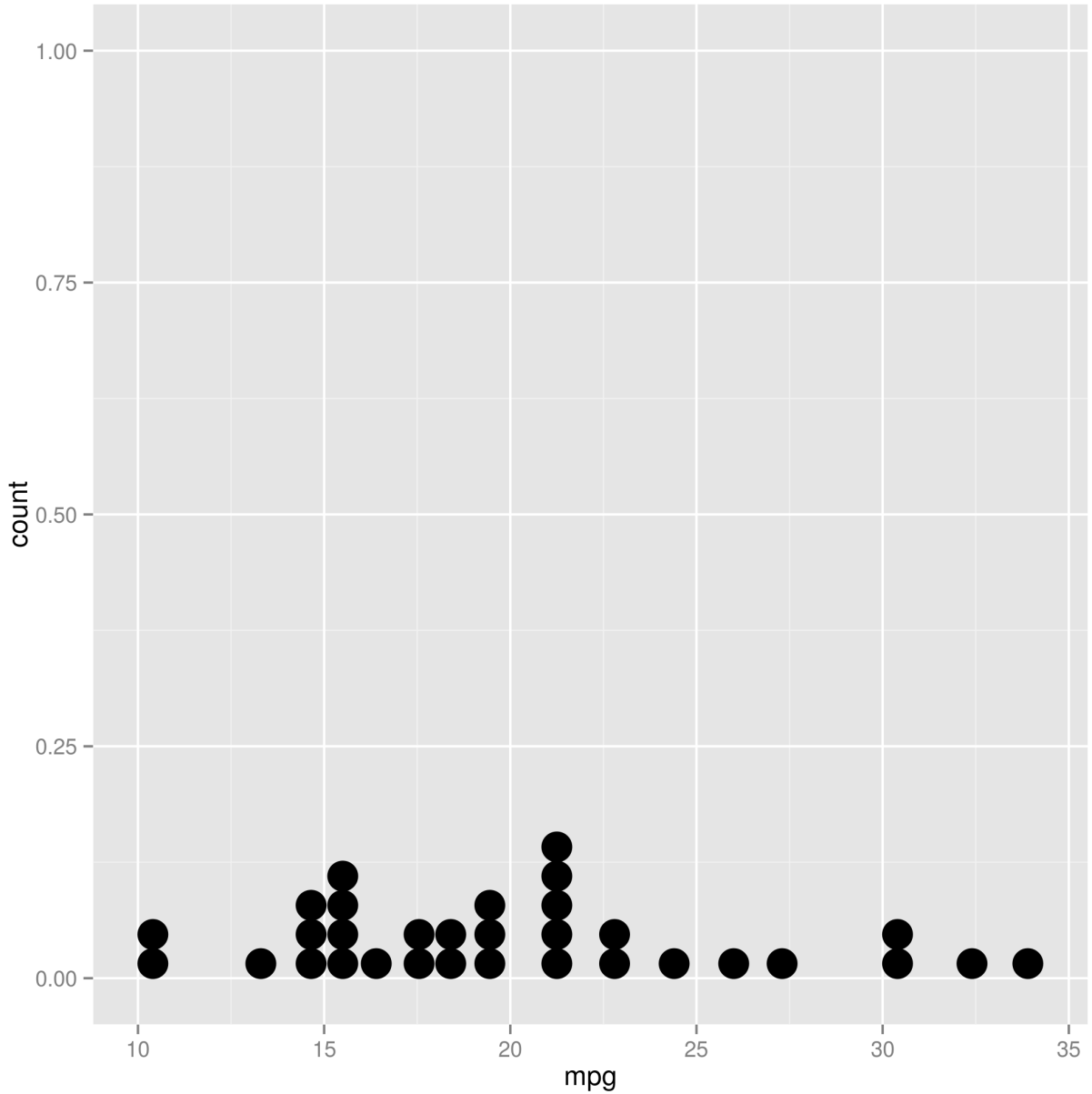
geom_dotplot()

- it sounds a bit like point, but the function is totally different
- infact it is a alternative for histograms for small sample sizes

```
> ggplot(mtcars, aes(x = mpg)) + geom_dotplot()
```

which results in:

dot plot example

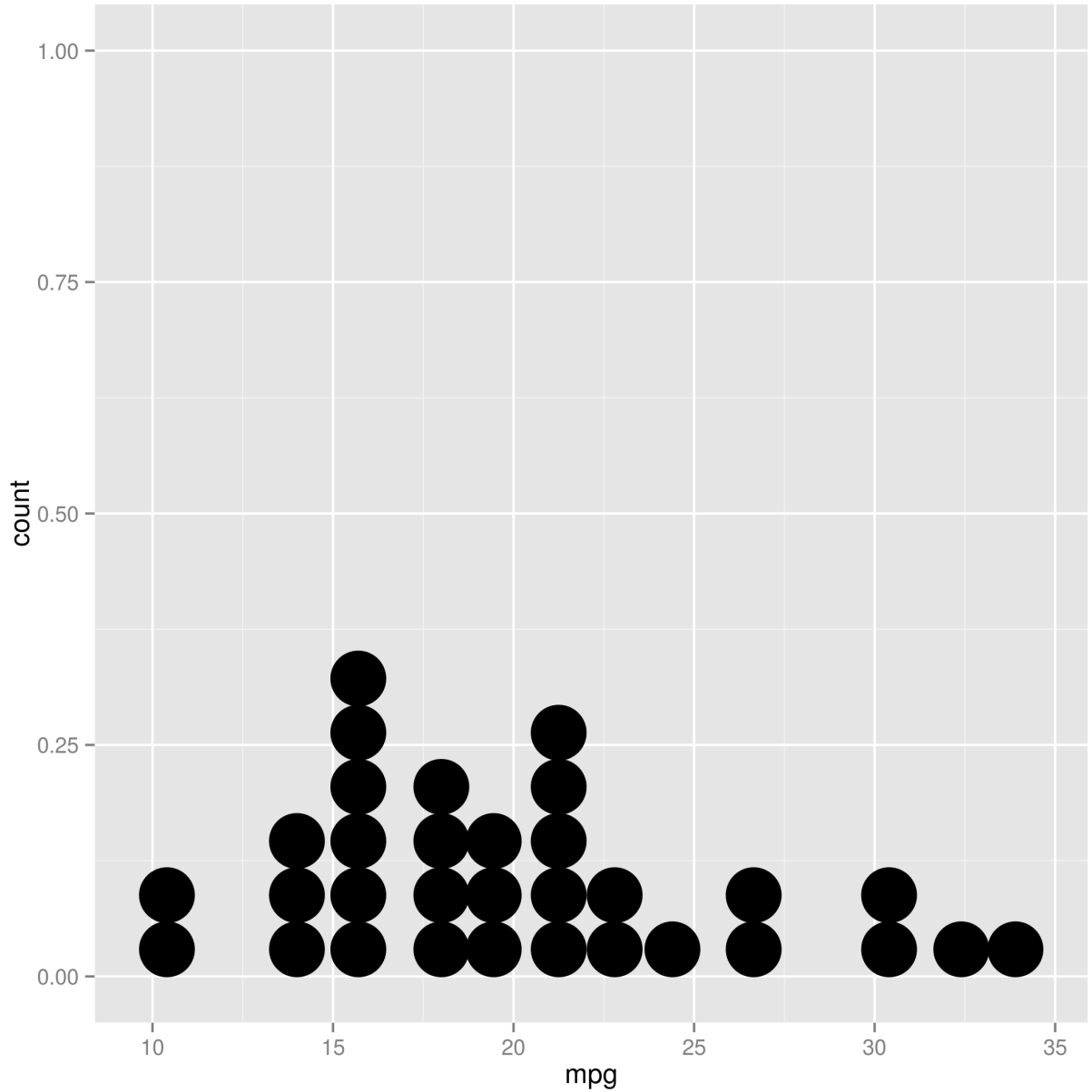


geom_dotplot()

```
> ggplot(mtcars, aes(x = mpg)) + geom_dotplot(binwidth = 1.5)
```

which results in:

dot plot example

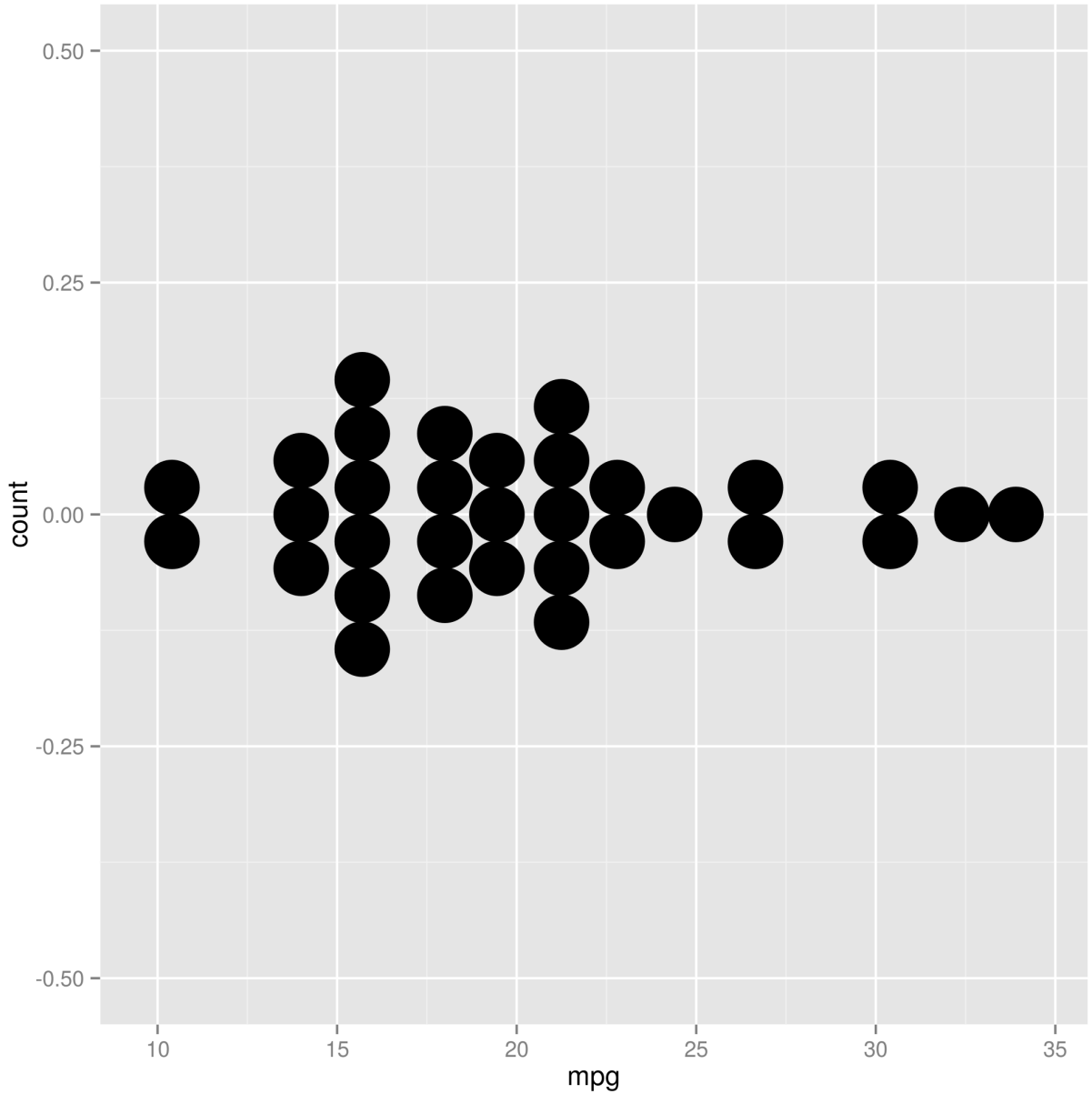


geom_dotplot()

```
> ggplot(mtcars, aes(x = mpg)) +
+     geom_dotplot(binwidth = 1.5, stackdir = "center")
```

which results in:

dot plot example



Exercises I

1. use again the GaltonFfamilies data set; produce a scatter plot of `childHeight` vs `midparentHeight`
2. add a trend line by using `geom_smooth()` without any arguments. Which method is used?
3. add a second trend line, this time a linear one!
4. now map the aesthetic `colour` to `gender` in the first line of the plot definition. What happens?

INBOthemes

- provides ggplot theme for elsevier and some nice color palettes
- new in Rstudio 0.99
  - data viewer displays more than 1000 lines
  - provides now filter and sort function
  - better completion
  - code diagnostics etc...
- density legends (OpenAnalytics)

What we've learned

- there are different tests of location, in general we can distinct
  - parametric and
  - non-parametric tests

What we've learned

- parametric test(s):
  - t-tests
- non-parametric test(s):
  - Wilcoxon tests

What we've learned

- if we talk about testing you should know the definition of the following terms, (a) for a one sample (b) for a two sample t-test
  - Null hypothesis
  - Alternative hypothesis
  - Test statistic
  - Significance level
  - Critical value
  - Decision rule
  - Type I error
  - Type II error
  - (Power)

Exercise

Lloyd et al (1969) studied the transferal of tritiated water (water containing a radioactive isotope of hydrogen with detrimental health effects) across the human chloroamnion (placental membrane). Permeability to tritium was measured for two groups of membranes, A and B. The data are

```
> data <- data.frame(values = c(0.8,0.83,1.89,1.04,1.45,
                                1.38,1.91,1.64,0.73,1.46,
                                1.15,0.88,0.9,0.74,1.21),
                     membrane = rep(LETTERS[1:2],c(10,5)))
```

1. check normality using histograms (and/or normal probability plots - `geom_qqplot()`)
2. Conduct a Wilcoxon rank sum test (`wilcox.test()`) for the alternative hypothesis that A is greater than B.

Exercise

To compare two different weight loss programs (X and Y) and to control of the confounding potential of genetics, 12 pairs of identical twins of similar weight were studied. Each paire of twins was randomly assigned to program X or Y and the amount of lost weight in pounds was recorded. Test if program X is superior to program Y.

1. State your hypothesis.
2. Do the appropriate test.
3. visualize.
4. State your conclusion.

Hint: The `granovaGG` package contains a function `granovagg.ds()` to visualize dependent samples. Try this plot and try to understand the plot elements.

Exercise

1. Install the `asbio` package and load the data set magnets using the `data()` command.
2. by typing the question mark followed by the name of the data set you get a description of the data

   ```
   > ?magnets
   ```

3. read this description. The question was *Was the pain reduction in the magnet group superior*. We assume a t-test is appropriate. Do all necessary steps to answer this question.