R Course,
AufLand Conference

Eduard Szöcs

26. November 2015

Institute for Environmental Sciences - University of Koblenz-Landau

Intro

Short Intro: Eduard Szöcs

- PhD student @Quantitative Landscape Ecology
- Environmental Sciences & Ecotoxicology
- Research:
  - Statistical Ecology - Eco(toxico)logical Statistics
  - Effects and distribution of pesticides in freshwaters
- R-Programming:
  - R-user for 6years
  - Author/Co-Author of 3 CRAN packages (taxize, webchem, rspear)
  - Other packages on github (restax, esmisc)
  - Minor contribitions to other pkgs (e.g. vegan)

`edild.github.io` @EduardSzoecs

Short Intro: Gunnar Oehmichen

Course Outline

- Short intro & course organisation, Software preparation
- An introduction to ggplot2
- Visualization & Exploration of models in R

Course material: `https://github.com/EDiLD/r_landau_2015`

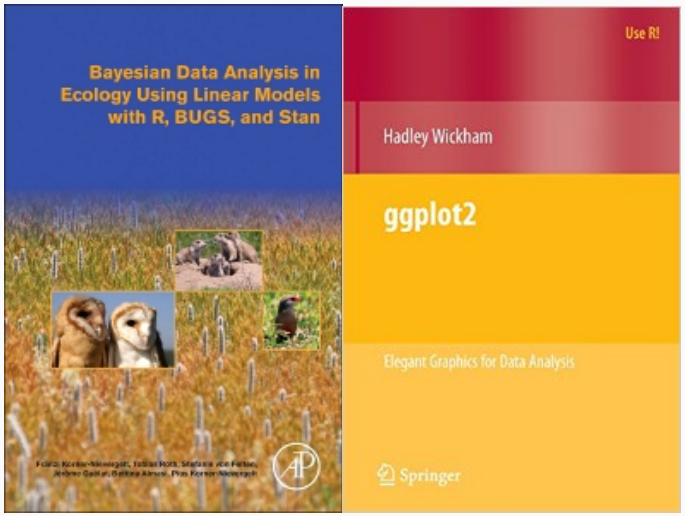
Preliminaries

- Download the course repository
- No formal R knowledge required to follow.
- Just open the ‘.R‘ files in RStudio and execute the script line by line: ‘CTRL + ENTER‘

Datasets

We will use 2 data sets in this course:

1. Frog monitoring in Swiss (Demo).
2. Diamond prices (Exercises).



FROGS ABUNDANCE

```
library(blmeco)
data(frogs)
head(frogs)
```

```
##   count1 count2 elevation year fish waterarea vegetation pondid      x
## 1     16     12       380 2013    0      2500          1 400301 649750
## 2      0      0       565 2009    0       300          1 400411 647350
## 3      0      0       430 2012    0       450          1 400603 650250
## 4      0      0       500 2012    0       348          1 400608 649400
## 5      0      0       450 2012    0       200          1 400701 646700
## 6      0      0       560 2010    0        42          1 400802 646500
##        y
## 1 248850
## 2 255750
## 3 244600
## 4 243850
## 5 240750
## 6 253650
```

```
frogs$fish <- factor(frogs$fish)
frogs$vegetation <- factor(frogs$vegetation)
```

Q: What influences frog abundance?

DIAMONDS DATASET

```
library(ggplot2)
data(diamonds)
head(diamonds)
```

```
##   carat       cut color clarity depth table price    x    y    z
## 1  0.23     Ideal     E     SI2  61.5    55   326 3.95 3.98 2.43
## 2  0.21   Premium     E     SI1  59.8    61   326 3.89 3.84 2.31
## 3  0.23      Good     E     VS1  56.9    65   327 4.05 4.07 2.31
## 4  0.29   Premium     I     VS2  62.4    58   334 4.20 4.23 2.63
## 5  0.31      Good     J     SI2  63.3    58   335 4.34 4.35 2.75
## 6  0.24 Very Good     J    VVS2  62.8    57   336 3.94 3.96 2.48
```

```
?diamonds
```

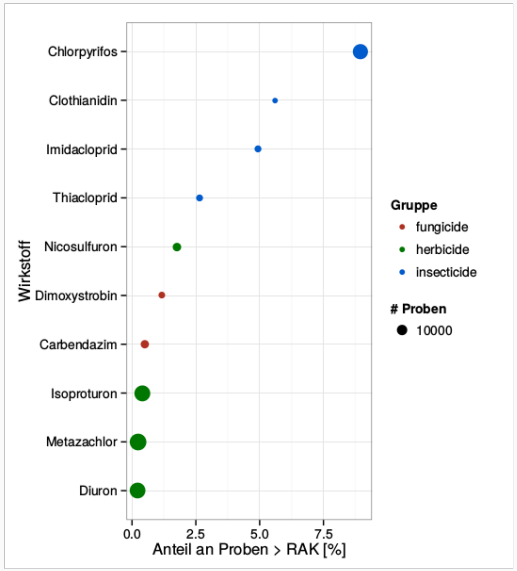
Q: What determines the price?

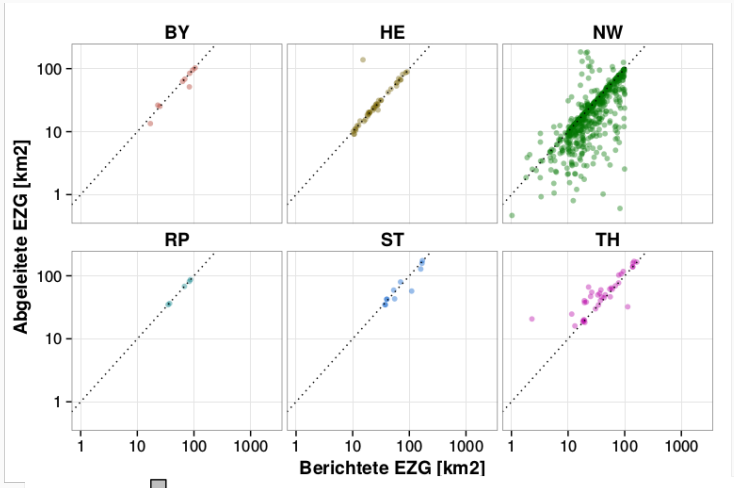
GGPLOT2

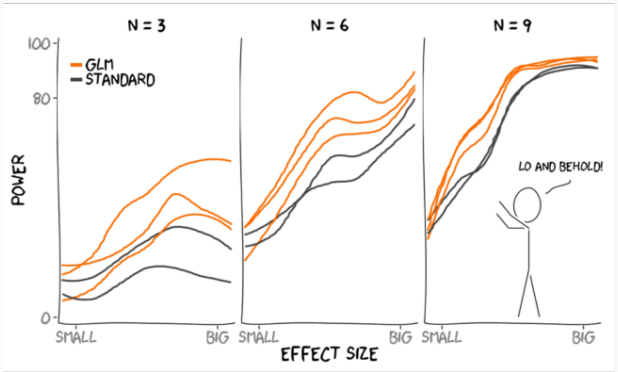
What is ggplot2?

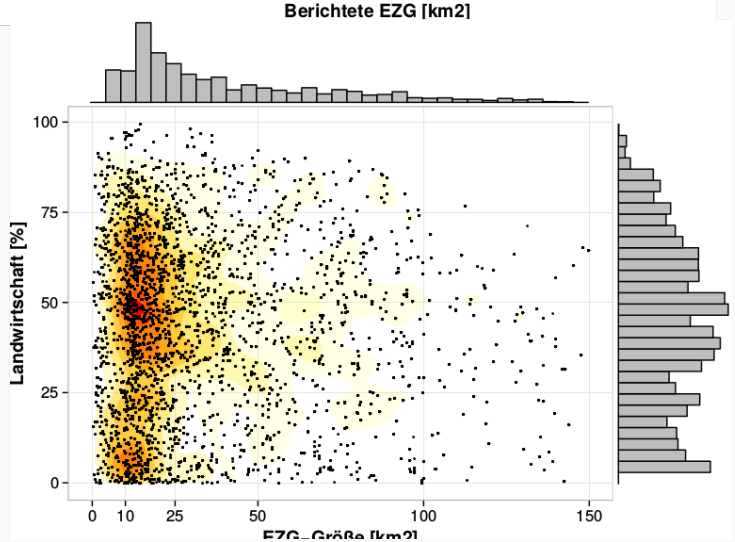
- A graphic system for R
- gg = **G**rammar of **G**raphics
- Grammar: Components that define a sentence
- ggplot defines a grammar to create plots
- Consistent, intuitive, easy to learn

Why ggplot2?


Chlorpyrifos
Clothianidin
Imidacloprid
Thiacloprid
Nicosulfuron
Dimoxystrobin
Carbendazim
Isoproturon
Metazachlor
Diuron
Wirkstoff
0.0
2.5
5.0
7.5
Anteil an Proben > RAK [%]
Gruppe
fungicide
herbicide
insecticide
Proben
10000

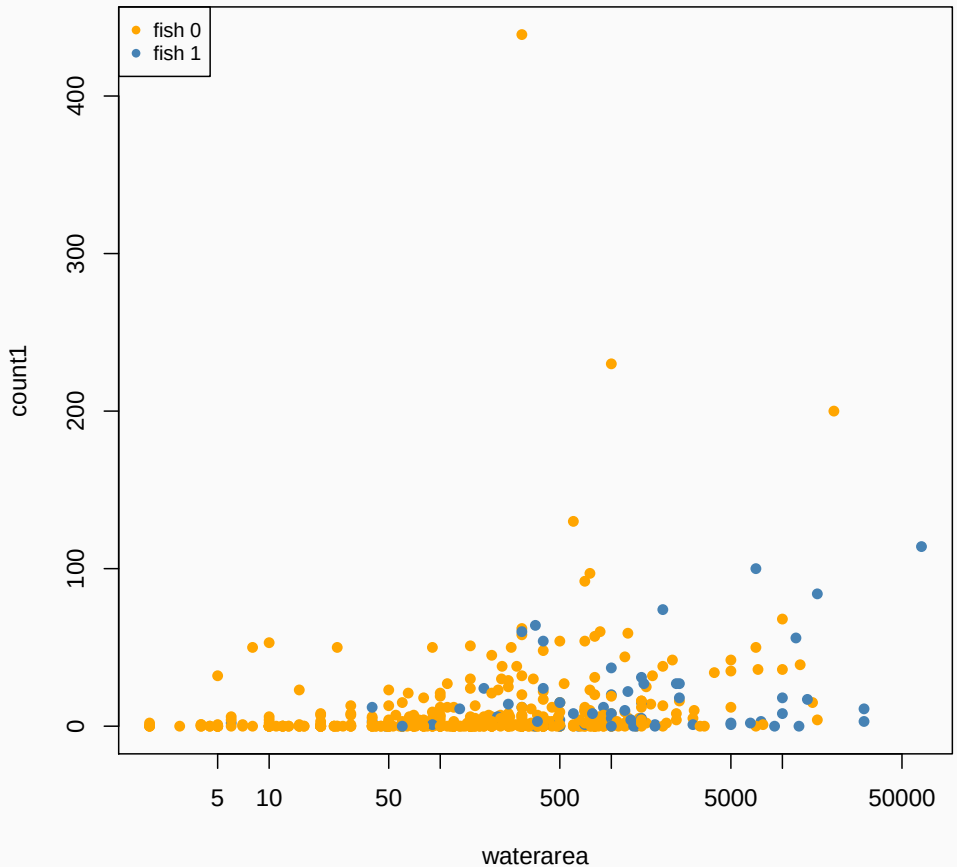

BY
HE
NW
RP
ST
TH
Abgeleitete EZG [km2]
Berichtete EZG [km2]
100
10
1
1000


N = 3
N = 6
N = 9
GLM
STANDARD
POWER
100
80
0
SMALL
BIG
EFFECT SIZE
LO AND BEHOLD!


Landwirtschaft [%]
100
75
50
25
0
0
10
25
50
100
150

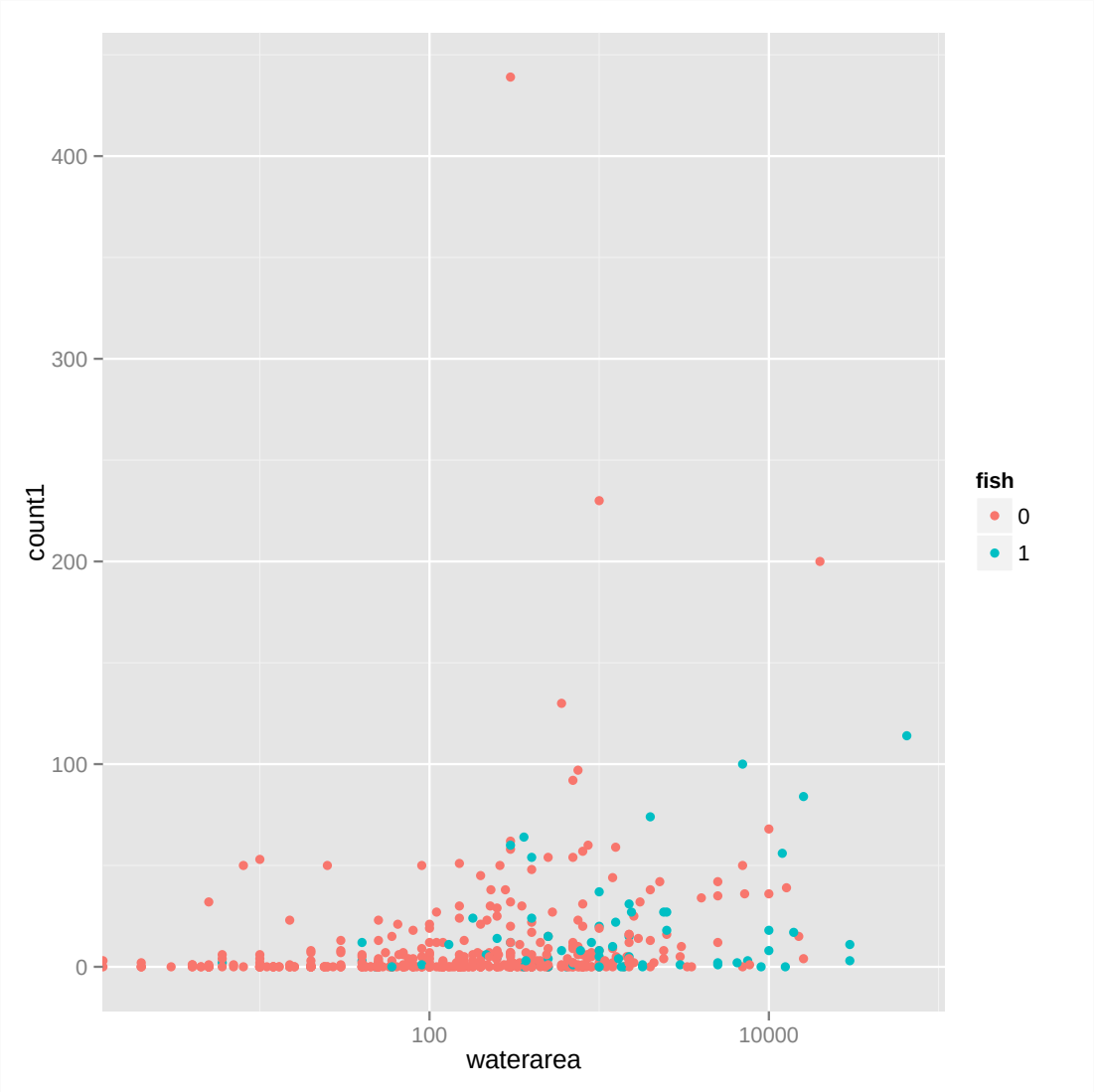
Why ggplot2?

```
cols <- c('orange', 'steelblue')
plot(count1 ~ waterarea, data = frogs, type = 'n', log = 'x')
with(frogs, points(waterarea, count1, col = cols[fish], pch = 16))
legend('topleft', legend = c('fish 0', 'fish 1'), pch = 16,
       col = cols,
       cex = 0.8)
```



Why ggplot2?

```
ggplot(frogs) +
  geom_point(aes(x = waterarea, y = count1, col = fish)) +
  scale_x_log10()
```



Why ggplot2?

base graphics

- more work
- legends, colors?
- ugly defaults
- multipanel plots?
- can do anything

ggplot2

- quick-and-dirty and complex
- Automatic legends, colors!
- nice defaults
- easy multipanel plots
- restrictions (e.g. 2nd y-axis)

Why ggplot2?

base graphics

- more work
- legends, colors?
- ugly defaults
- multipanel plots?
- can do anything

ggplot2

- quick-and-dirty and complex
- Automatic legends, colors!
- nice defaults
- easy multipanel plots
- restrictions (e.g. 2nd y-axis)

Tip 1:

If a plot is too much work to draw with ggplot2, reconsider if it's a good representation of your data.

Terminology

```
ggplot(frogs) +
  geom_point(
    aes(x = waterarea, y = count1, col = fish)) +
  scale_x_log10()
```

ggplot() The main function. Can specify the data set and variables **globally**.

geom A geometric object: `geom_point, geom_line, geom_text, geom_violin`

aes aesthetics: maps a variable to the properties of a geom: shape, color, fill, linetype, transparency (alpha)

scale How are data mapped visually (log, continuous, discrete, date, colors, sizes)

Terminology

```
ggplot(frogs) +
  geom_point(
    aes(x = waterarea, y = count1, col = fish)) +
  scale_x_log10()
```

ggplot() The main function. Can specify the data set and variables **globally**.

geom A geometric object: `geom_point, geom_line, geom_text, geom_violin`

aes aesthetics: maps a variable to the properties of a geom: shape, color, fill, linetype, transparency (alpha)

scale How are data mapped visually (log, continuous, discrete, date, colors, sizes)

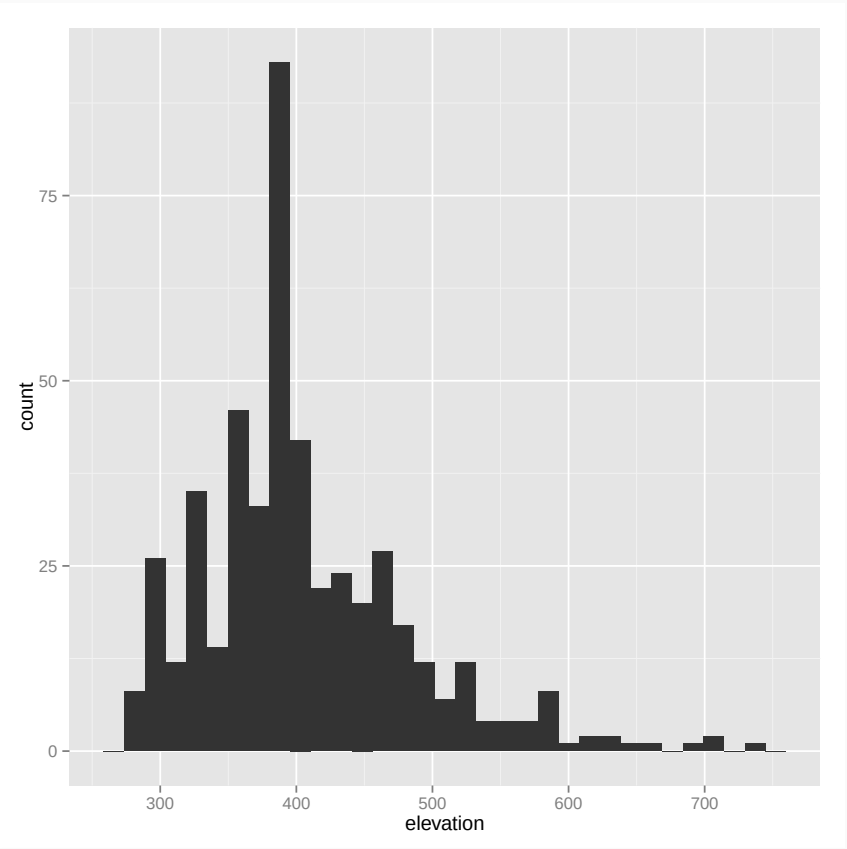
Tip 2:

Never use `qplot()` (=quick plot)! - You won't learn the grammar...

Univariate Data

geom: Histogram

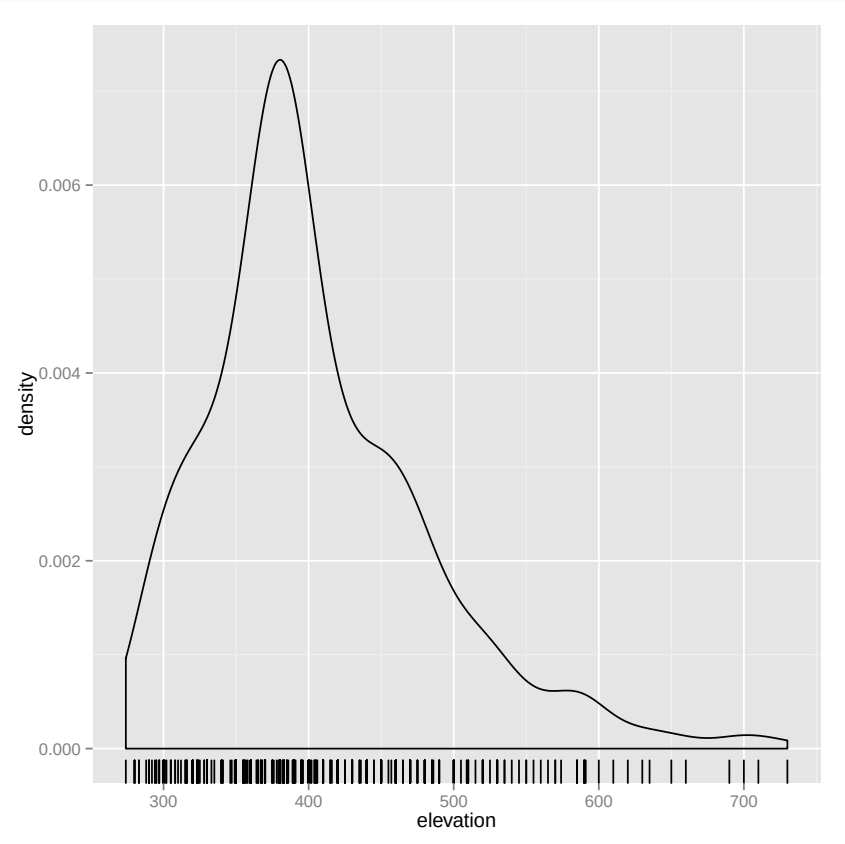
```
ggplot(frogs) +          # use the frogs dataset
  geom_histogram(        # display a histogramm
    aes(x = elevation)   # take the variable 'elevation' from the dataset
    )


## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust
this.
```



geom: Density + rug

We can add multiple geoms of the same variable to the plot

```
ggplot(frogs, aes(x = elevation)) +    # plot the 'elevation' from the frogs data
  geom_density() +                        # display a density
  geom_rug()                              # display a rug
```

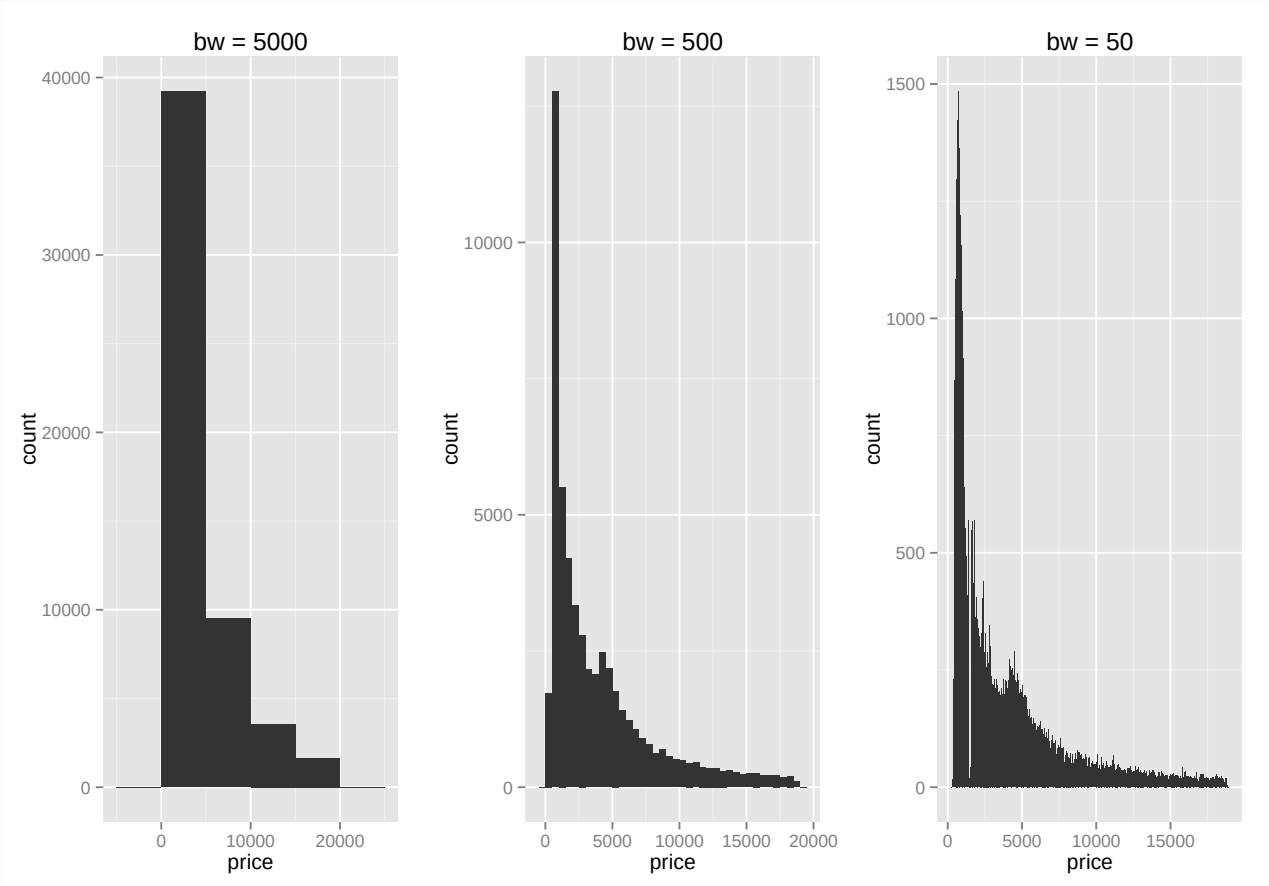


Exercise 1:

Plot a histogramm of diamond prices at different binwidths (50, 500, 5000).

How does this affect the plot?

Histogram

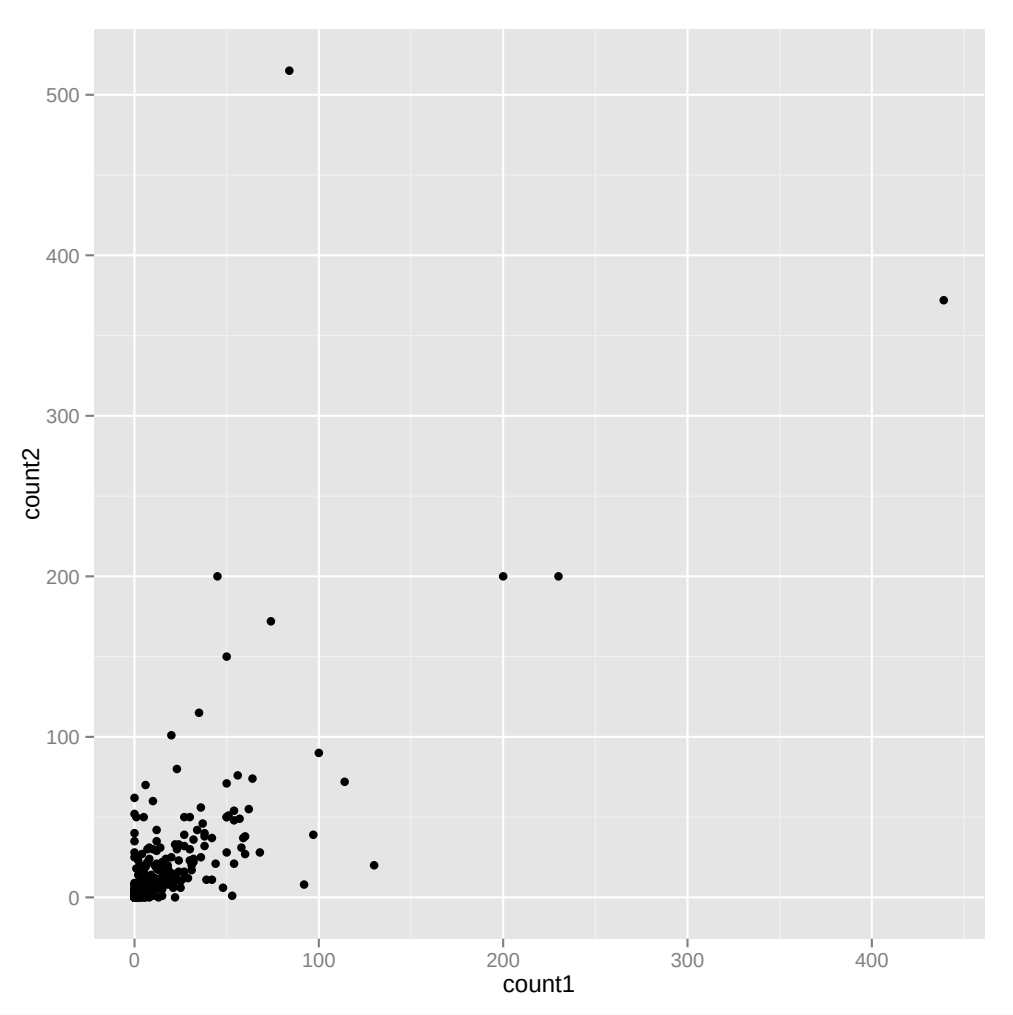

bw = 5000
bw = 500
bw = 50
count
price
0
10000
20000
30000
40000
5000
15000
1000
1500
500

Bivariate data

geom: Point (=Cont. x Cont.)

The most basic plot.

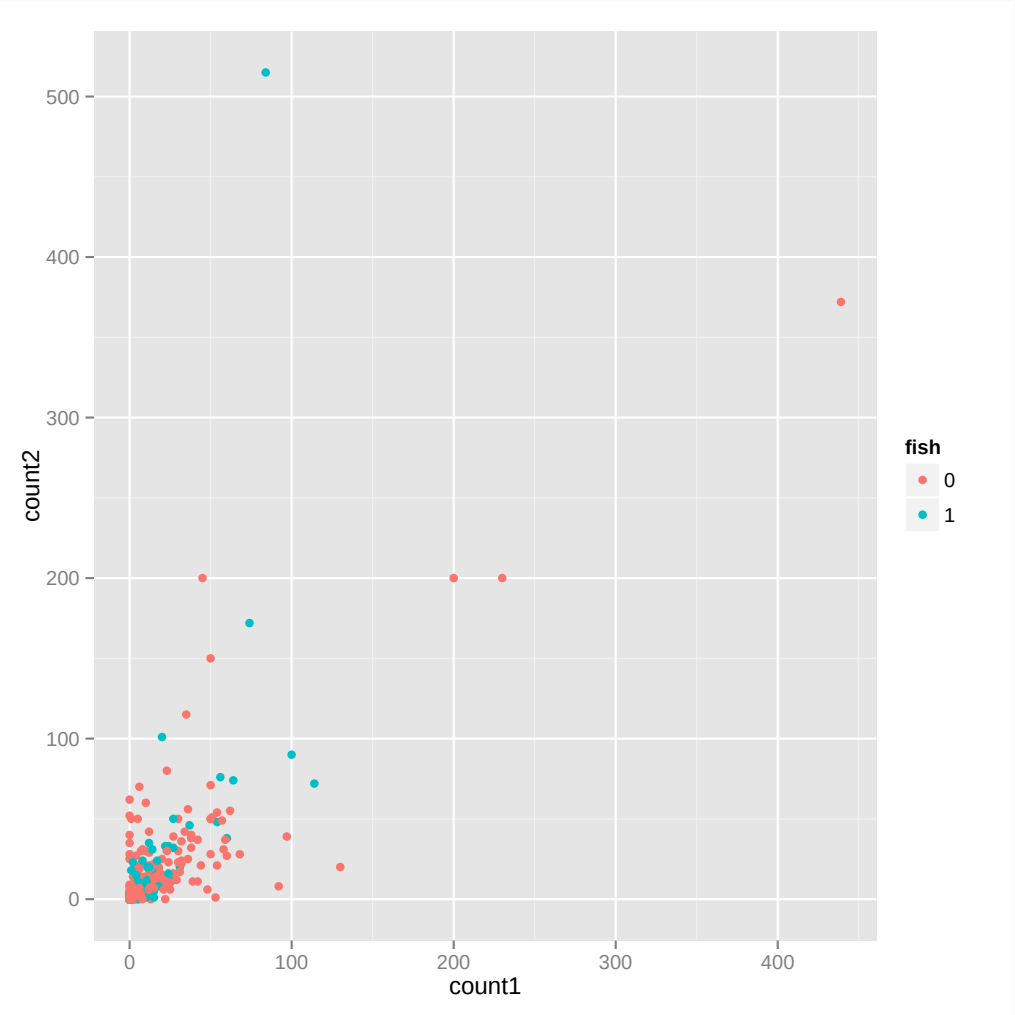
```
ggplot(frogs) +
  geom_point(aes(x = count1, y = count2))
```



aesthetics: Add color.

Differentiate between fish and no-fish:

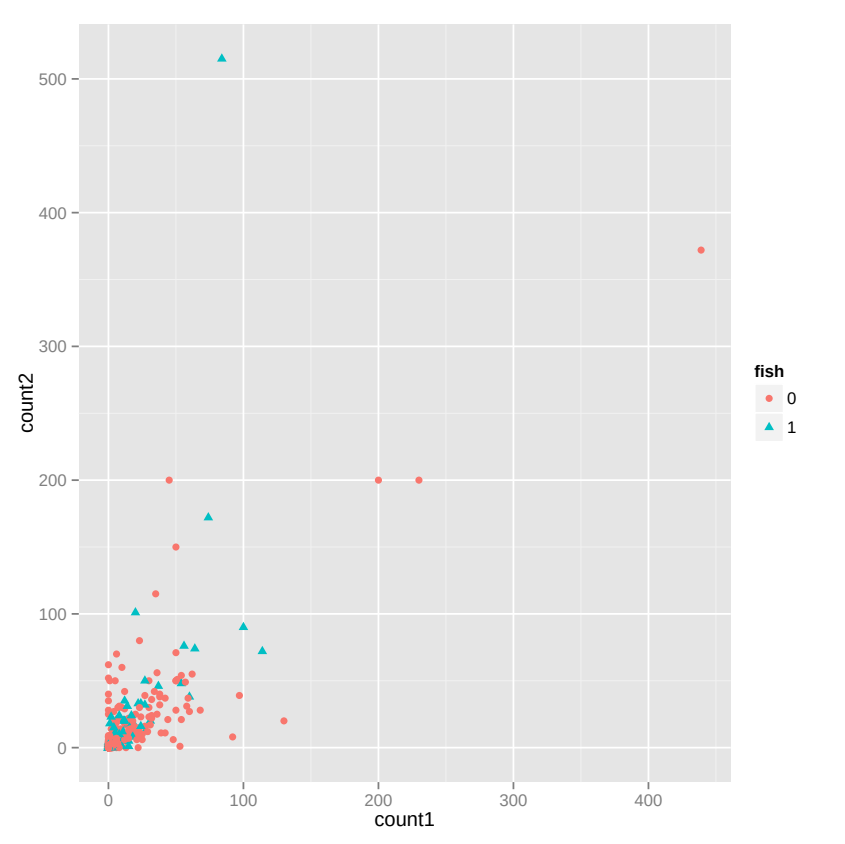
```
ggplot(frogs) +
  geom_point(aes(x = count1, y = count2, col = fish))
```



AESTHETICS: COLORS + SHAPE

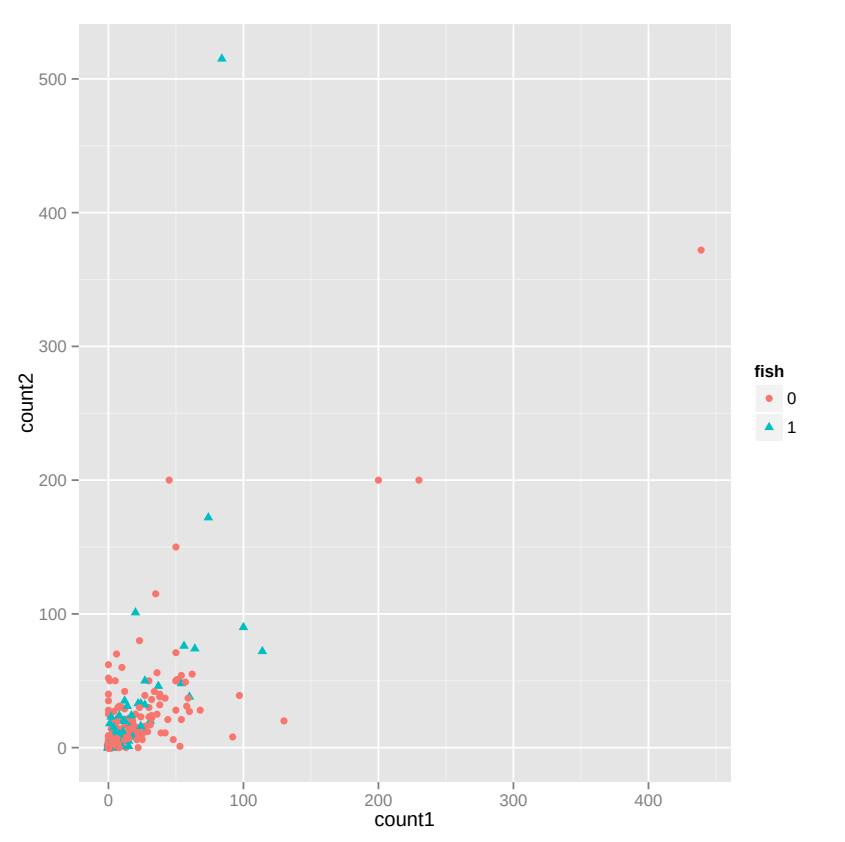
Can also differentiate by shape:

```
ggplot(frogs) +
  geom_point(aes(x = count1, y = count2, col = fish, shape = fish))
```



Can also differentiate by shape:

```
ggplot(frogs) +
  geom_point(aes(x = count1, y = count2, col = fish, shape = fish))
```

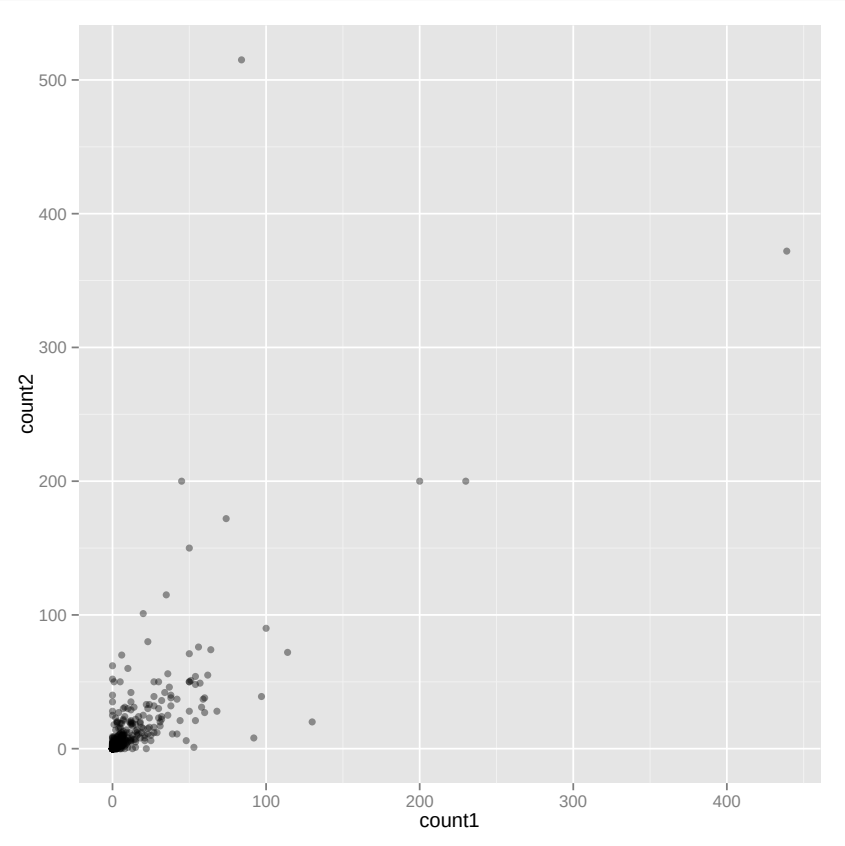


Tip 3:

Do not use redundant aesthetics.

AESTHETICS: TRANSPARENCY

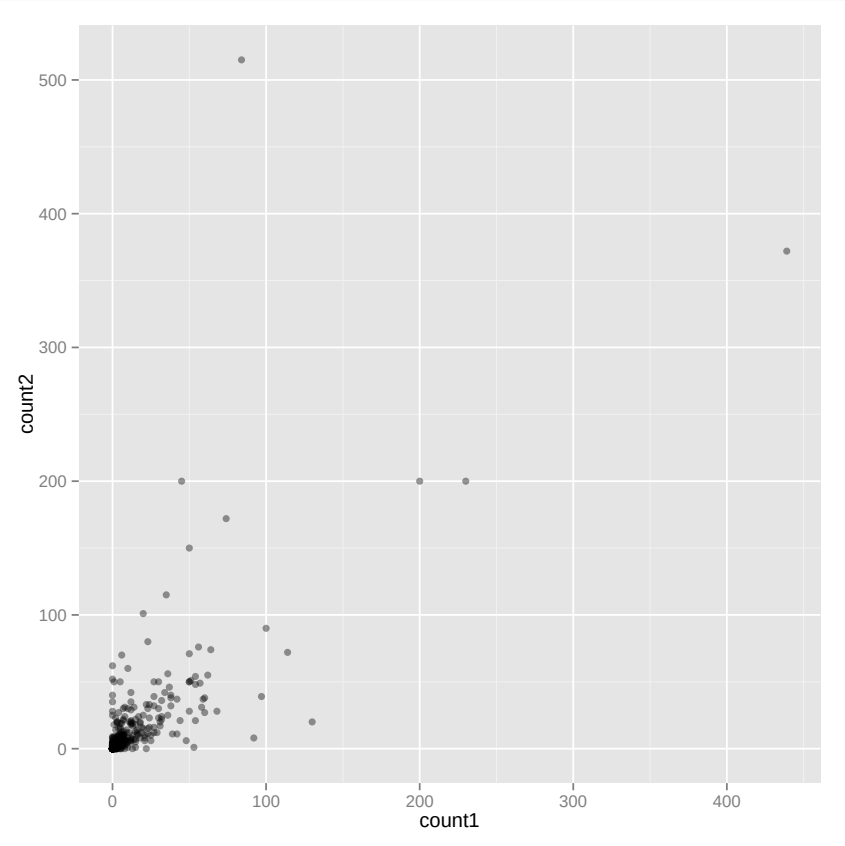
```
ggplot(frogs) +
  geom_point(aes(x = count1, y = count2), alpha = 0.4)
```



Q: alpha is not within aes(), why?

AESTHETICS: TRANSPARENCY

```
ggplot(frogs) +
  geom_point(aes(x = count1, y = count2), alpha = 0.4)
```

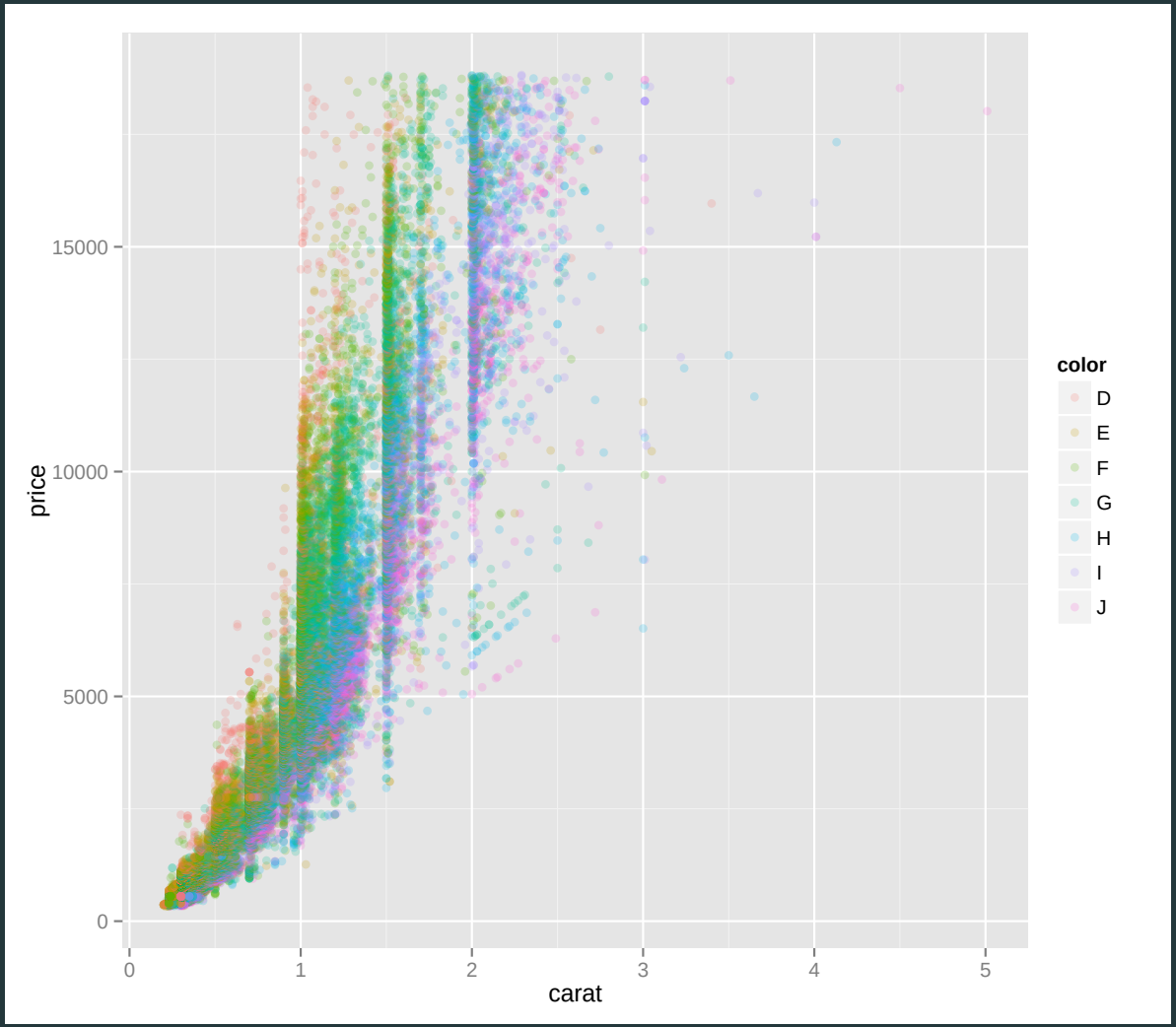


Q: alpha is not within aes(), why?

Tip 4:

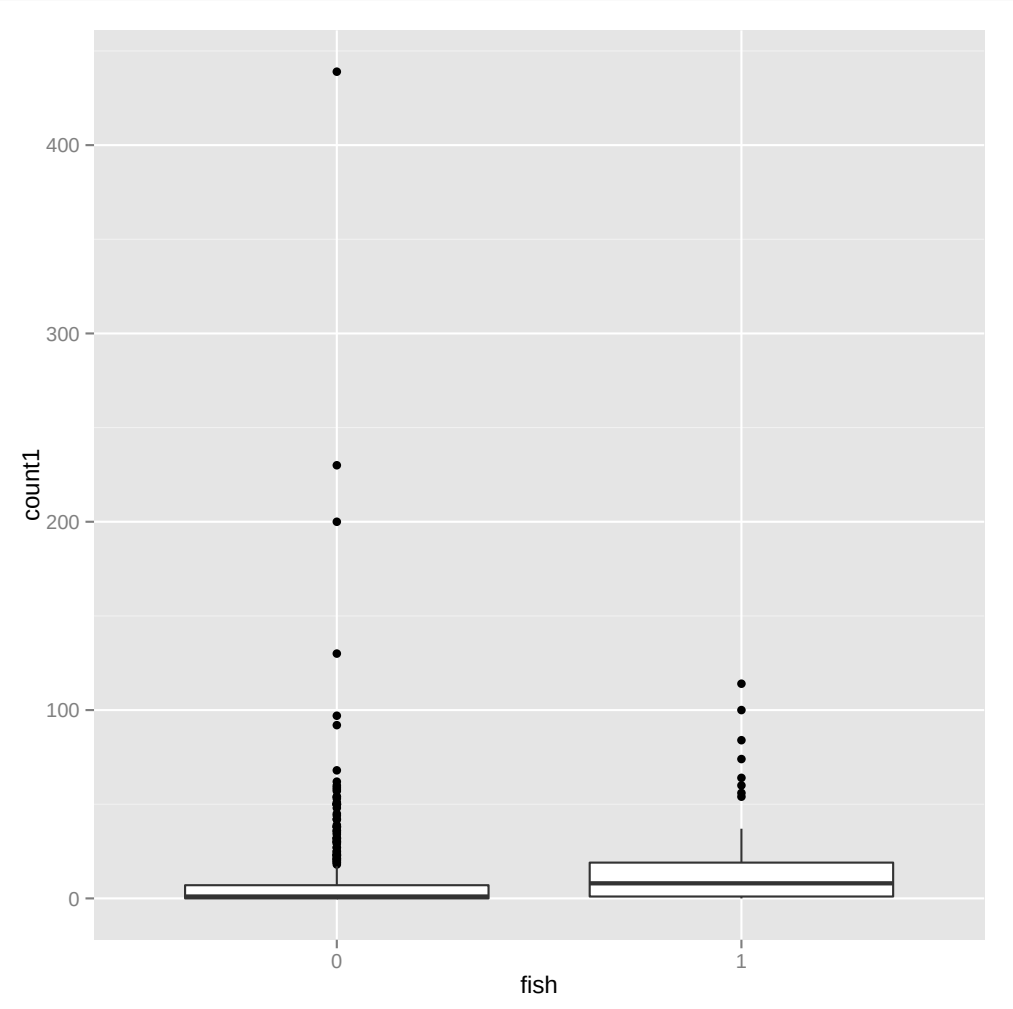
Transparency is very powerful for lots of data!

Exercise 2:
Create this plot:



geom: boxplot (Discrete x Continuous)

```
ggplot(frogs) +
  geom_boxplot(aes(x = fish, y = count1))
```



Highly-skewed distribution - not much information visibile

SCALES: LOG SCALE

```
ggplot(frogs) +
  geom_boxplot(aes(x = fish, y = count1 + 1)) +
  scale_y_log10()
```