

# R COURSE, AUFLAND CONFERENCE

---

Eduard Szöcs

26. November 2015

Institute for Environmental Sciences - University of Koblenz-Landau




## INTRO

---

- 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 contributions to other pkgs (e.g. vegan)

[edild.github.io](https://edild.github.io)

 [@EduardSzoecs](https://twitter.com/EduardSzoecs)



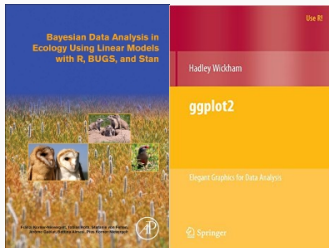
- Short intro & course organisation, Software preparation
- An introduction to ggplot2
- Visualization & Exploration of models in R

Course material: [https://github.com/EDiLD/r\\_landau\\_2015](https://github.com/EDiLD/r_landau_2015)

- 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'

We will use 2 data sets in this course:

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



## FROGS ABUNDANCE

```
library(blmeeco)
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?



```
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
```

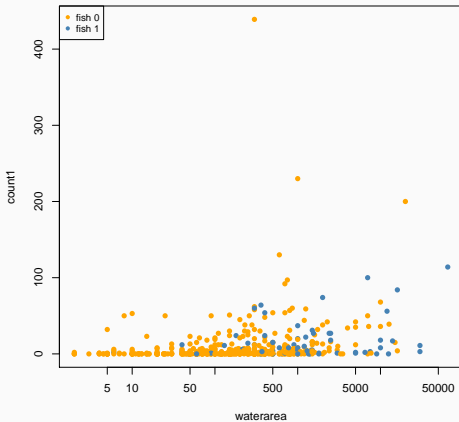
Q: What determines the price?

GGPLOT2

- 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

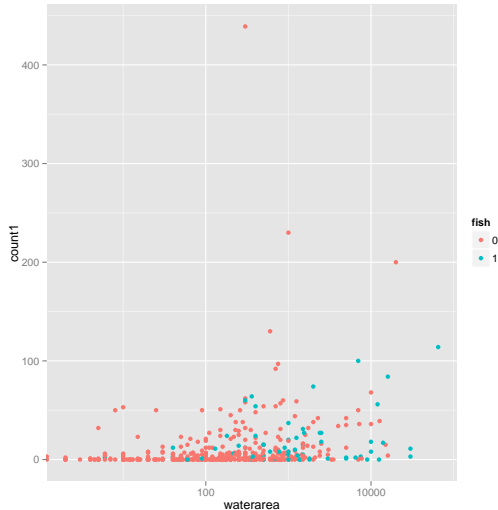
## WHY GGPLOT2?

```
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 GGLOT2?

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



## 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)

## 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.

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

**ggplot()** The main function. Can specify the dataset and variables for all geoms here.

**geom** A geometric object: **geom\_point**, **geom\_line**,  
**geom\_text**, **geom\_violin**

**aes** aesthetics: shape, transparency (alpha), color, fill, linetype.

**scale** How is the data displayed (log, continuous, discrete, date)



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

**ggplot()** The main function. Can specify the dataset and variables for all geoms here.

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

**aes** aesthetics: shape, transparency (alpha), color, fill, linetype.

**scale** How is the data displayed (log, continuous, discrete, date)

## Tip 2:

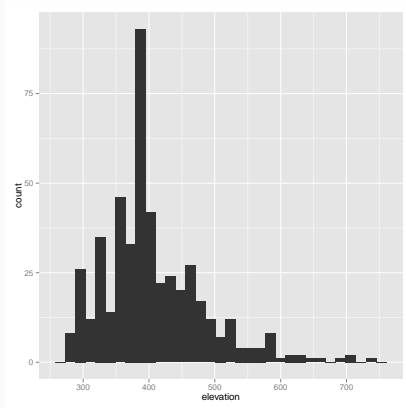
Never use `qplot()` (=quick plot)! - You won't learn the grammar...

## UNIVARIATE DATA

# HISTOGRAM

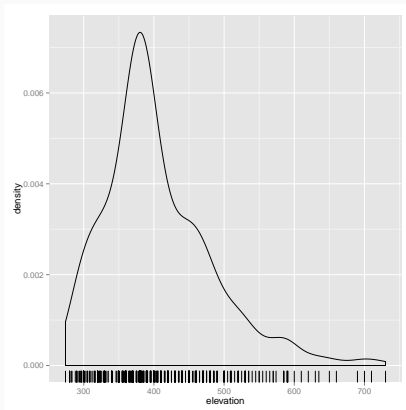
```
ggplot(frogs) +           # use the frogs dataset
  geom_histogram(         # display a histogram
    aes(x = elevation)    # take the variable 'elevation' from the dataset
  )
```

## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.



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
```

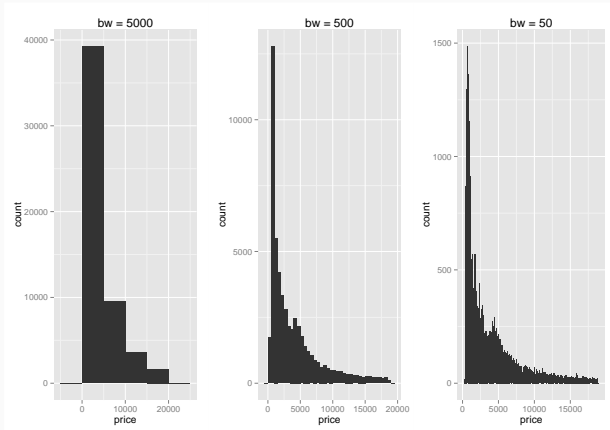


EXERCISE 1:

PLOT A HISTOGRAMM OF DIAMOND PRICES AT DIFFERENT BINWIDTHS  
(50, 500, 5000).

HOW DOES THIS AFFECT THE PLOT?

# HISTOGRAM



## BIVARIATE DATA

The most basic plot.

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

