

Script example (R Markdwon)

Pedro J. Aphalo

30 March 2015

Contents

1	knitr setup	1
2	Data	1
3	Data preparation	2
4	EDA (exploratory data analysis)}	2
5	Test of significance	6

1 knitr setup

```
library(knitr)
```

2 Data

We use cork production data from package `agridat`.

```
library(agridat)
data("box.cork")

summary(box.cork)
```

```
##      tree      N      E      S
## Min.   : 1.00  Min.   :30.00  Min.   :29.00  Min.   : 27.00
## 1st Qu.: 7.75  1st Qu.:38.50  1st Qu.:35.00  1st Qu.: 34.00
## Median :14.50  Median :46.50  Median :41.50  Median : 43.00
## Mean   :14.50  Mean   :50.54  Mean   :46.18  Mean   : 49.68
## 3rd Qu.:21.25  3rd Qu.:60.00  3rd Qu.:54.25  3rd Qu.: 66.25
## Max.   :28.00  Max.   :91.00  Max.   :80.00  Max.   :100.00
##      W
## Min.   :25.00
## 1st Qu.:35.00
## Median :41.50
## Mean   :45.18
## 3rd Qu.:58.00
## Max.   :77.00
```

3 Data preparation

We melt from 4 columns into a single column.

```
library(reshape2)
cork.melted <- melt(box.cork,
  id.vars = "tree",
  variable.name = "trunk.side",
  value.name = "cork.cg")
```

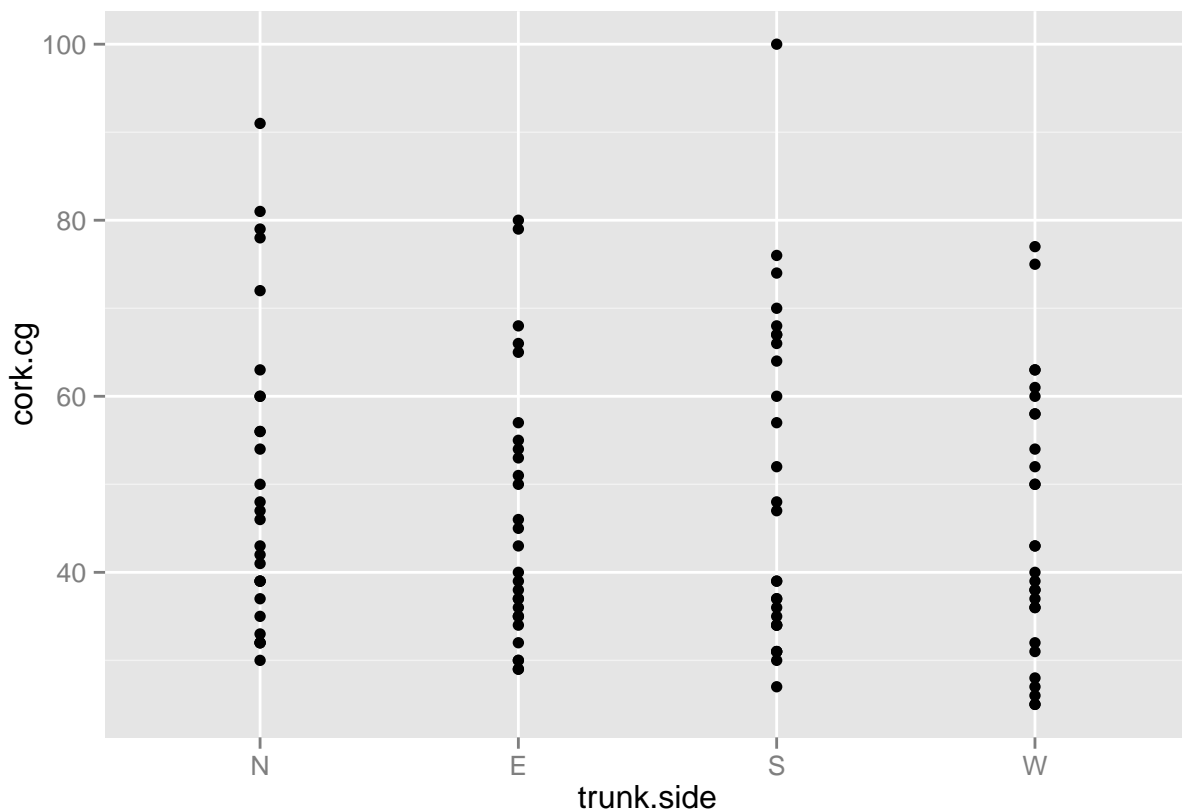
4 EDA (exploratory data analysis)}

We do exploratory plotting with 'ggplot2'.

```
library(ggplot2)
```

First you save a simple *dot plot*.

```
fig.cork <- ggplot(data = cork.melted,
  aes(x = trunk.side, y = cork.cg)) +
  geom_point()
fig.cork
```

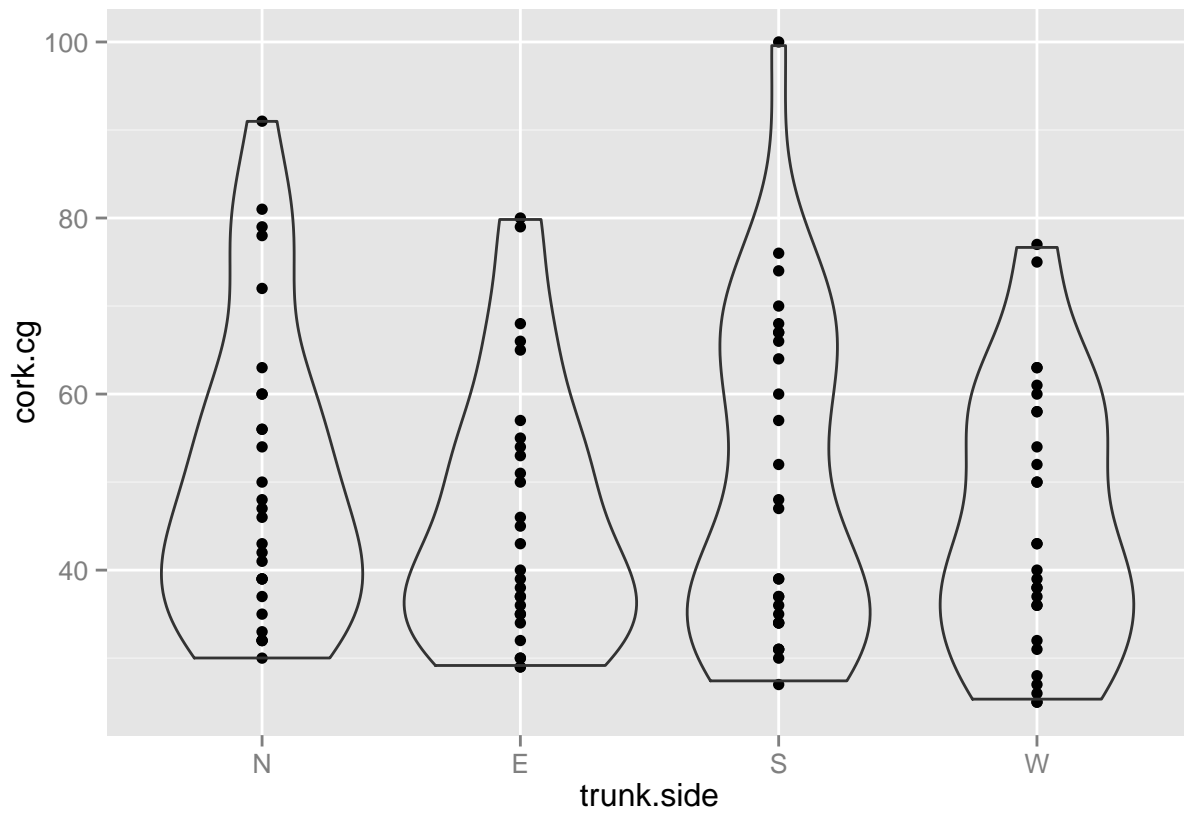


We add to the plot saved above *boxplots*, and also change the *y* scale to logarithmic.

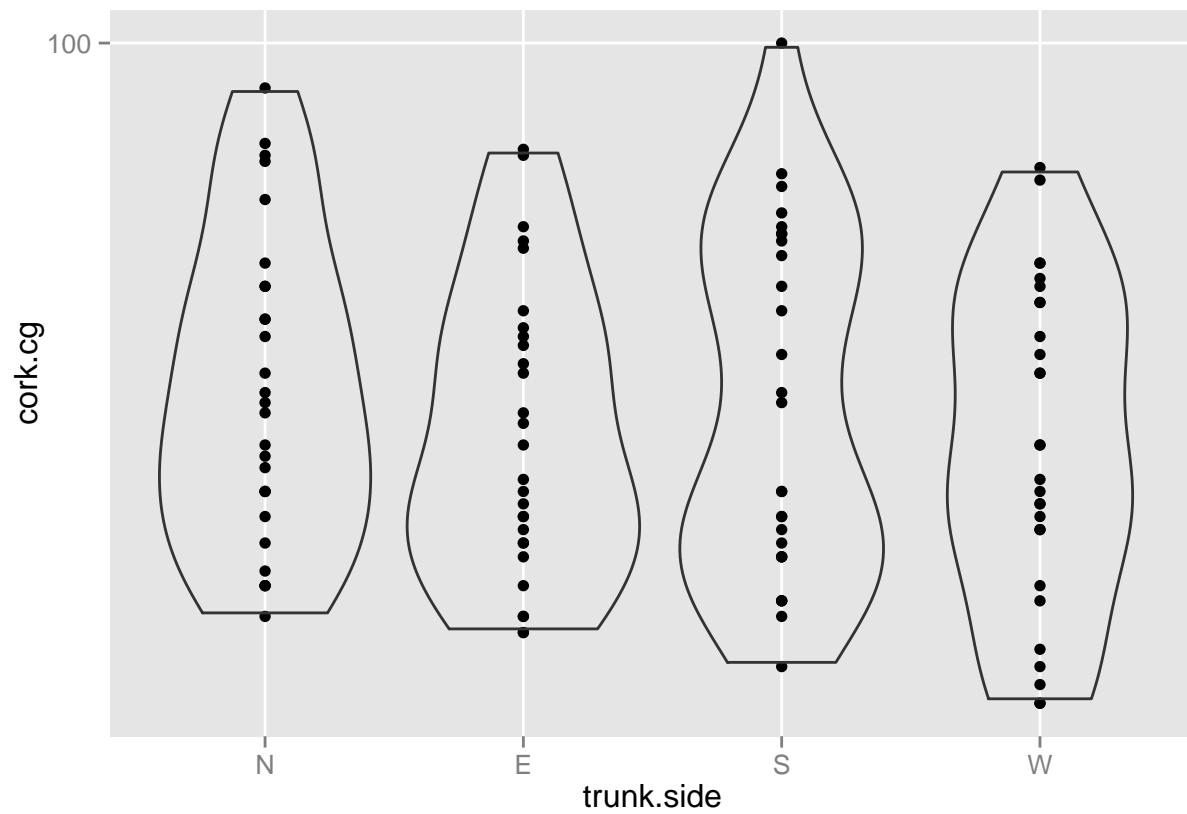
```
<<>>= fig.cork + geom_boxplot(fill = NA) fig.cork + geom_boxplot(fill = NA) + scale_y_log10() @
```

We add to the plot saved above *violin plots*, and also change the *y* scale to logarithmic.

```
fig.cork + geom_violin(fill = NA)
```

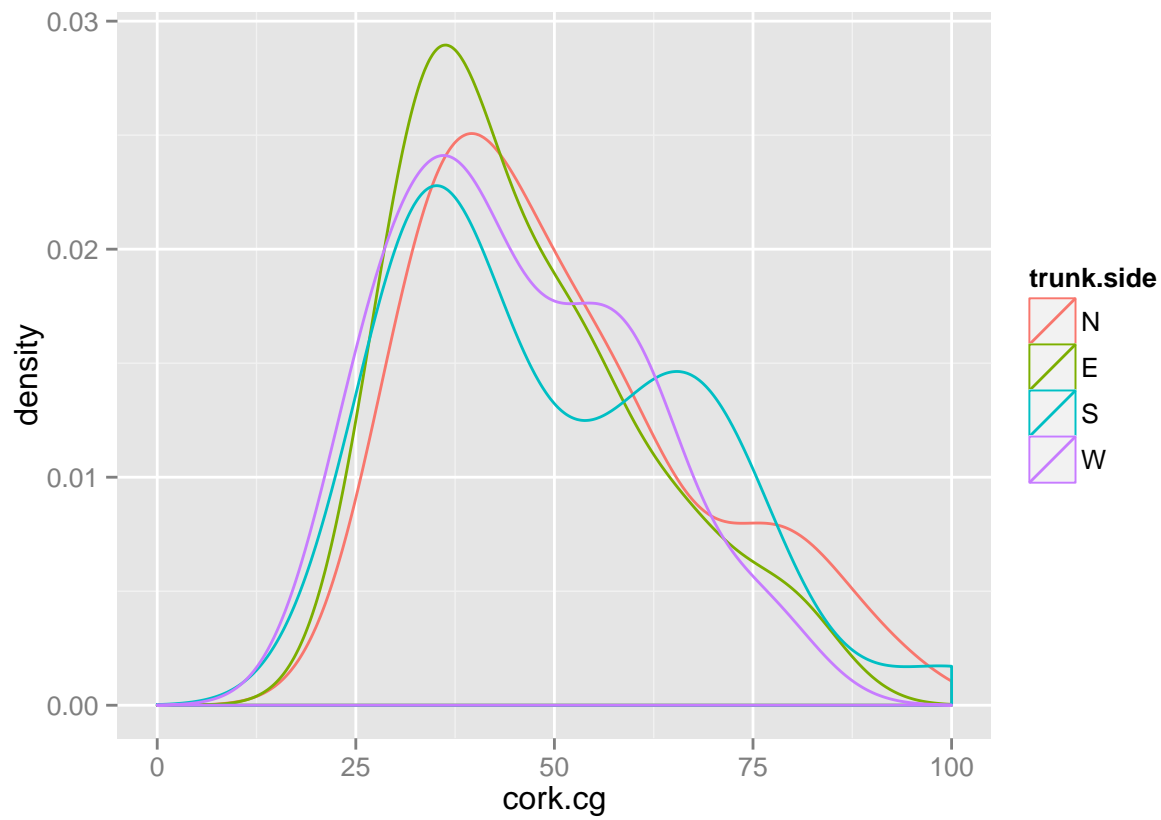


```
fig.cork + geom_violin(fill = NA) + scale_y_log10()
```



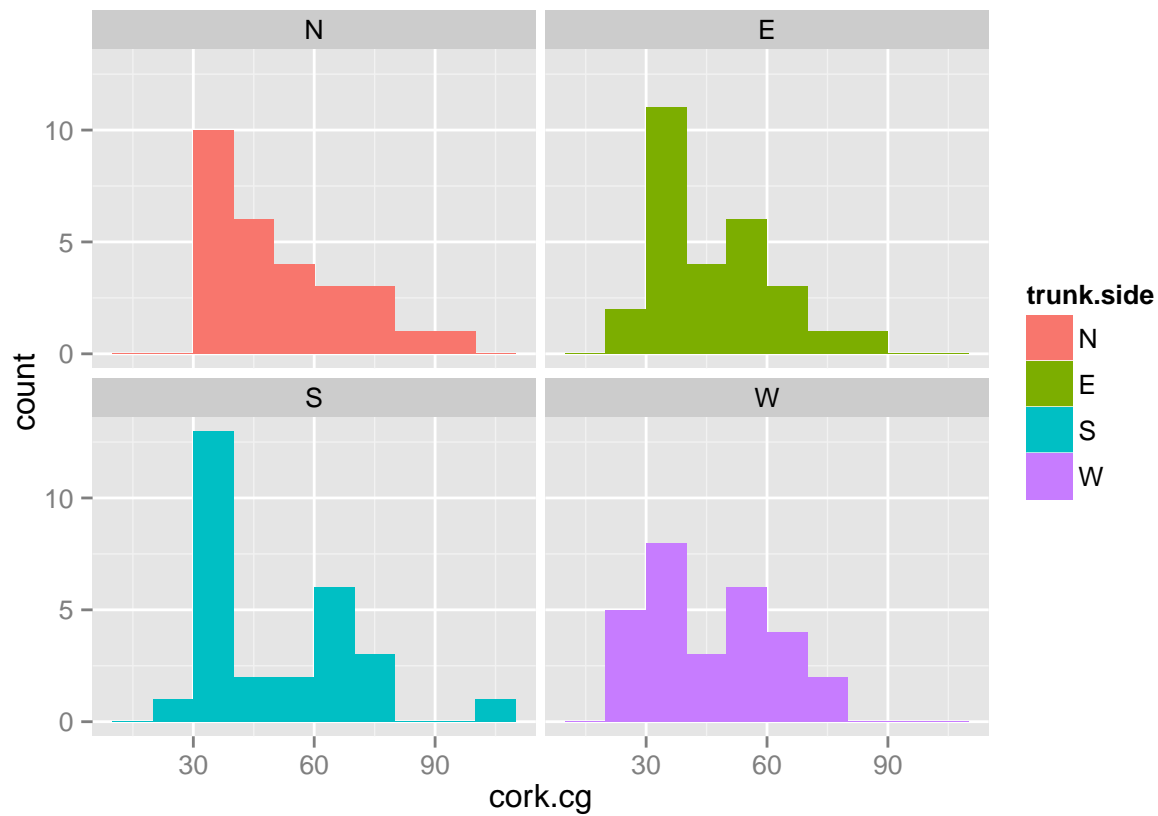
We create and save a new plot, with fitted densities to assess the shape of the distributions and their differences.

```
fig.cork1 <- ggplot(data = cork.melted,
                    aes(x = cork.cg, colour = trunk.side)) +
  geom_density() + xlim(0, NA)
fig.cork1
```



Here we use histograms instead.

```
fig.cork2 <- ggplot(data = cork.melted, aes(x = cork.cg, fill = trunk.side)) +
  geom_histogram(binwidth = 10) +
  facet_wrap(~trunk.side, ncol = 2)
fig.cork2
```



5 Test of significance

We fit a linear model, calculate an ANOVA table and check diagnosis plots

```
cork.fit <- lm(log10(cork.cg) ~ trunk.side + tree, data = cork.melted)
anova(cork.fit)
```

```
## Analysis of Variance Table
##
## Response: log10(cork.cg)
##          Df Sum Sq Mean Sq F value Pr(>F)
## trunk.side  3  0.04264  0.0142147   0.6786  0.5670
## tree        1  0.00010  0.0001031   0.0049  0.9442
## Residuals 107  2.24123  0.0209460
```

```
plot(cork.fit)
```

