Notes on using R

Pedro J. Aphalo

Git: tag ‘none’, committed with hash ‘none’ on ‘none’

Contents

Preface

This series of Notes cover different aspects of the use of R. They are meant to be use as a complement to a course or book, as explanations are short and terse. We do not discuss here statistics, just R as a tool and language for data manipulation and display.

1 Plots with ggplot2, ggrepel and ggpmisc

1.1 Packages used in this chapter

For executing the examples listed in this chapter you need first to load the following packages from the library:

```
library(ggplot2)
library(ggrepel)
library(ggpmisc)
```

We set a font larger size than the default

```
theme_set(theme_grey(16))
```

1.2 ggpmisc

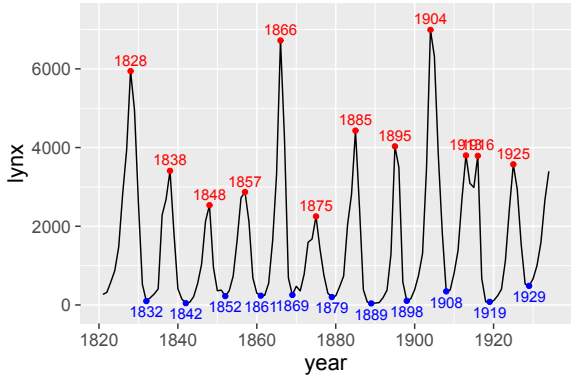
1.2.1 New stats

Package ggpmisc provides new stats: `stat_peaks()`, `stat_valleys()`, and `stat_poly_eq()`. Peaks and valleys are local (or global) maxima and minima. These stats return the x and y values at the peaks or valleys plus suitable labels, and default aesthetics that make easy their use with several different geoms, including `geom_point`, `geom_text`, `geom_label`, `geom_vline`, `geom_hline` and `geom_rug`, and also with geoms defined by package ggrepel. Some examples follow.

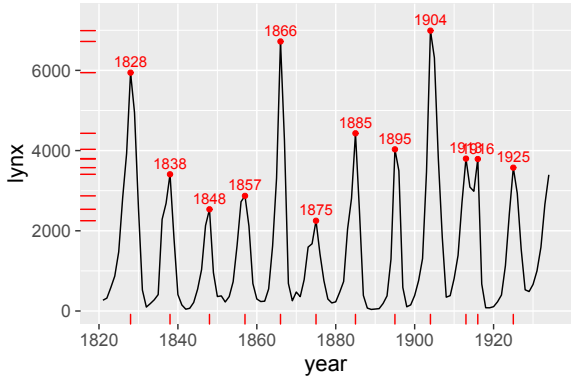
1.2.2 Peaks and valleys

```
lynx.df <- data.frame(year = as.numeric(time(lynx)), lynx = as.matrix(lynx))
```

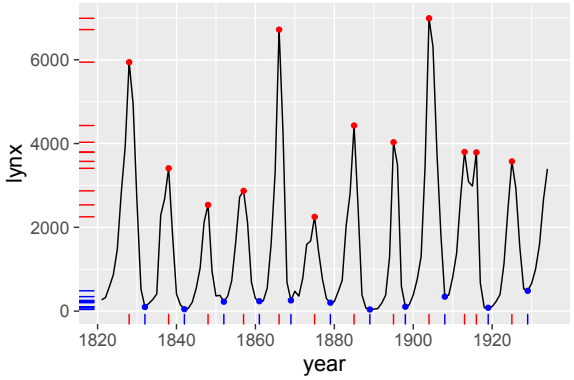
```
ggplot(lynx.df, aes(year, lynx)) + geom_line() +
  stat_peaks(colour = "red") +
  stat_peaks(geom = "text", colour = "red",
             vjust = -0.5, x.label.fmt = "%4.0f") +
  stat_valleys(colour = "blue") +
  stat_valleys(geom = "text", colour = "blue",
               vjust = 1.5, x.label.fmt = "%4.0f") +
  ylim(-100, 7300)
```



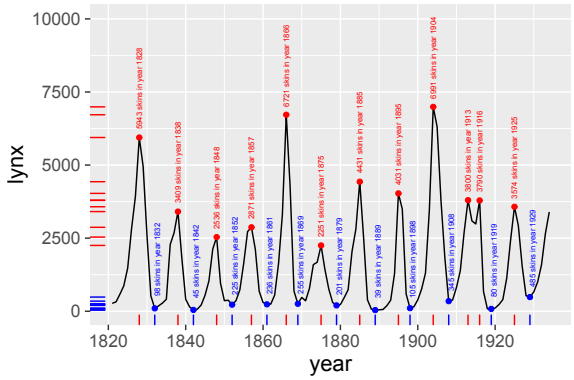

```
ggplot(lynx.df, aes(year, lynx)) + geom_line() +
  stat_peaks(colour = "red") +
  stat_peaks(geom = "rug", colour = "red") +
  stat_peaks(geom = "text", colour = "red",
             vjust = -0.5, x.label.fmt = "%4.0f") +
  ylim(NA, 7300)
```



```
ggplot(lynx.df, aes(year, lynx)) + geom_line() +
  stat_peaks(colour = "red") +
  stat_peaks(geom = "rug", colour = "red") +
  stat_valleys(colour = "blue") +
  stat_valleys(geom = "rug", colour = "blue")
```




```
ggplot(lynx.df, aes(year, lynx)) + geom_line() +
  stat_peaks(colour = "red") +
  stat_peaks(geom = "rug", colour = "red") +
  stat_peaks(geom = "text", colour = "red",
             hjust = -0.1, label.fmt = "%4.0f",
             angle = 90, size = rel(2),
             aes(label = paste(..y.label..,
                               "skins in year", ..x.label..))) +
  stat_valleys(colour = "blue") +
  stat_valleys(geom = "rug", colour = "blue") +
  stat_valleys(geom = "text", colour = "blue",
             hjust = -0.1, vjust = 1, label.fmt = "%4.0f",
             angle = 90, size = rel(2),
             aes(label = paste(..y.label..,
                               "skins in year", ..x.label..))) +
  ylim(NA, 10000)
```



Of course, if one finds use for it, the peaks and/or valleys can be plotted on their own.

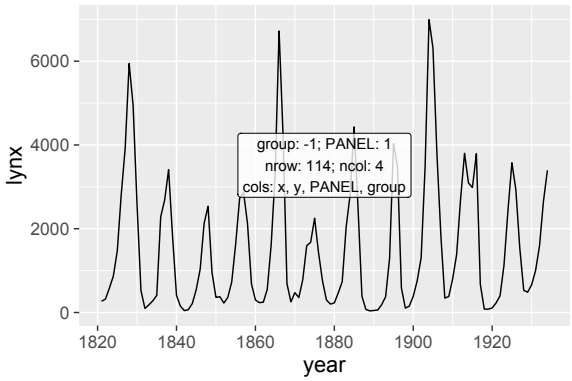
```
ggplot(lynx.df, aes(year, lynx)) +
  stat_peaks(geom = "line") + stat_valleys(geom = "line")
```

1.2.3 Learning and/or debugging

A very simple stat named `stat_debug()` can save the work of adding print statements to the code of stats to get information about what data is being passed to the `compute_group()` function. Because the code of this function is stored in a `ggproto` object, at the moment it is impossible to directly set breakpoints in it. This stat may also help users diagnose problems with the mapping of aesthetics in their code or just get a better idea of how the internals of ggplot2 work.

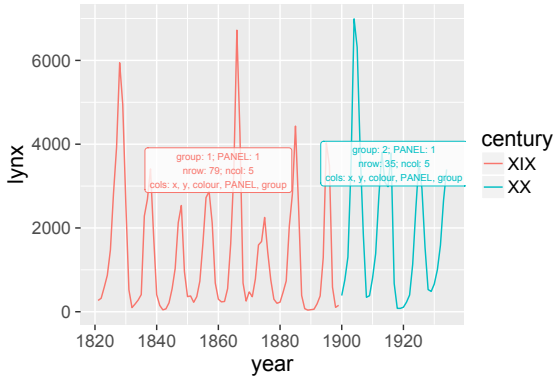
```
ggplot(lynx.df, aes(year, lynx)) + geom_line() +
  stat_debug(alpha = 0.8)

##        x    y nrow ncol           colnames group
## 1 1877.5 3515  114    4 x, y, PANEL, group    -1
##   PANEL
## 1     1
```




```
lynx.df$century <- ifelse(lynx.df$year >= 1900, "XX", "XIX")
ggplot(lynx.df, aes(year, lynx, color = century)) +
  geom_line() +
  stat_debug(alpha = 0.8, size = rel(2.5))

##      x    y nrow ncol                   colnames
## 1 1860 3380   79    5 x, y, colour, PANEL, group
##   group PANEL
## 1     1     1
##      x      y nrow ncol
## 1 1917 3535.5   35    5
##                    colnames group PANEL
## 1 x, y, colour, PANEL, group     2     1
```



1.3 ggrepel

1.3.1 New geoms

Package ggrepel provides two new geoms: `geom_text_repel` and `geom_label_repel`. They are used similarly to `geom_text` and `geom_label` but the text or labels "repel" each other so that they rarely overlap unless the plot is very crowded.

```
try(detach(package:ggpmisc))
try(detach(package:ggrepel))
try(detach(package:ggplot2))
```

2 Further reading about R

2.1 Introductory texts

2.2 Texts on specific aspects

2.3 Advanced texts

Bibliography

Chang, W. (2013). *R Graphics Cookbook*. 1-2. Sebastopol: O'Reilly Media, p. 413. ISBN: 9781449316952. URL: http://medcontent.metapress.com/index/A65RM03P4874243N.pdf.

Dalgaard, P. (2008). *Introductory Statistics with R*. Springer, p. 380. ISBN: 0387790543.

Everitt, B. and T. Hothorn (2011). *An Introduction to Applied Multivariate Analysis with R*. Springer, p. 288. ISBN: 1441996494. URL: http://www.amazon.co.uk/Introduction-Applied-Multivariate-Analysis-Use/dp/1441996494.

Faraway, J. J. (2004). *Linear Models with R*. Boca Raton, FL: Chapman & Hall/CRC, p. 240. URL: http://www.maths.bath.ac.uk/~jjf23/LMR/.

Faraway, J. J. (2006). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman & Hall/CRC Taylor & Francis Group, p. 345. ISBN: 158488424X.

Fox, J. (2002). *An {R} and {S-Plus} Companion to Applied Regression*. Thousand Oaks, CA, USA: Sage Publications. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion/index.html.

Fox, J. and H. S. Weisberg (2010). *An R Companion to Applied Regression*. SAGE Publications, Inc, p. 472. ISBN: 141297514X. URL: http://www.amazon.com/An-R-Companion-Applied-Regression/dp/141297514X.

Ihaka, R. and R. Gentleman (1996). 'R: A Language for Data Analysis and Graphics'. In: *J. Comput. Graph. Stat.* 5, pp. 299–314.

Matloff, N. (2011). *The Art of R Programming: A Tour of Statistical Software Design*. No Starch Press, p. 400. ISBN: 1593273843. URL: http://www.amazon.com/The-Art-Programming-Statistical-Software/dp/1593273843.

Murrell, P. (2011). *R Graphics, Second Edition (Chapman & Hall/CRC The R Series)*. CRC Press, p. 546. ISBN: 1439831769. URL: http://www.amazon.com/Graphics-Second-Edition-Chapman-Series/dp/1439831769.

Pinheiro, J. C. and D. M. Bates (2000). *Mixed-Effects Models in S and S-Plus*. New York: Springer.

Teetor, P. (2011). *R Cookbook*. 1st ed. Sebastopol: O'Reilly Media, p. 436. ISBN: 9780596809157.

Wickham, H. (2014). *Advanced R*. Chapman & Hall/CRC The R Series. CRC Press. ISBN: 9781466586970. URL: https://books.google.fi/books?id=G5PNBQAAQBAJ.

– (2015). *R Packages*. O'Reilly Media. ISBN: 9781491910542. URL: https://books.google.fi/books?id=eqOxBwAAQBAJ.

Xie, Y. (2013). *Dynamic Documents with R and knitr (Chapman & Hall/CRC The R Series)*. Chapman and Hall/CRC, p. 216. ISBN: 1482203537. URL: http://www.amazon.com/Dynamic-Documents-knitr-Chapman-Series/dp/1482203537.

Zuur, A. F., E. N. Ieno and E. Meesters (2009). *A Beginner's Guide to R*. 1st ed. Springer, p. 236. ISBN: 0387938362. URL: http://www.amazon.com/Beginners-Guide-Use-Alain-Zuur/dp/0387938362.