Finding my way around R

ggplot2: Barplots

MARCH 17, 2009

**One R Tip A Day (http://onertipaday.blogspot.com/2009/01/statistical-visualizations-part-2.html)** uses R basic graphics to visualise migration to the United States during 1820-2006.
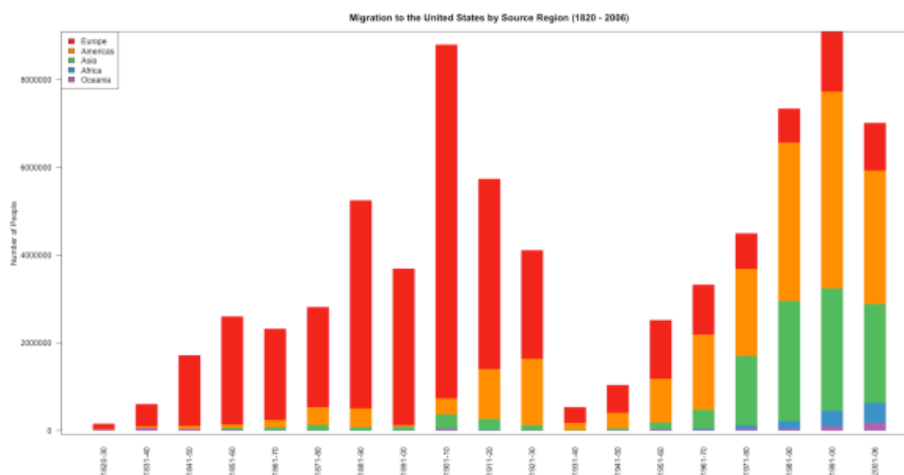
Again, as usual, let's reproduce this in **ggplot2 (http://had.co.nz/ggplot2)**.

First, load ggplot2 and dataset

```
> library(ggplot2)
> df <- structure(c(106487, 495681, 1597442,
    2452577, 2065141, 2271925, 4735484, 3555352,
    8056040, 4321887, 2463194, 347566, 621147,
    1325727, 1123492, 800368, 761550, 1359737,
    1073726, 36, 53, 141, 41538, 64759, 124160,
    69942, 74862, 323543, 247236, 112059, 16595,
    37028, 153249, 427642, 1588178, 2738157,
    2795672, 2265696, 11951, 33424, 62469,
    74720, 166607, 404044, 426967, 38972, 361888,
    1143671, 1516716, 160037, 354804, 996944,
    1716374, 1982735, 3615225, 4486806, 3037122,
    17, 54, 55, 210, 312, 358, 857, 350, 7368,
    8443, 6286, 1750, 7367, 14092, 28954, 80779,
    176893, 354939, 446792, 33333, 69911, 53144,
    29169, 18005, 11704, 13363, 18028, 46547,
    14574, 8954, 2483, 14693, 25467, 25215,
    41254, 46237, 98263, 185986), .Dim = c(19,
    5), .Dimnames = list(c("1820-30", "1831-40",
    "1841-50", "1851-60", "1861-70", "1871-80",
    "1881-90", "1891-00", "1901-10", "1911-20",
    "1921-30", "1931-40", "1941-50", "1951-60",
    "1961-70", "1971-80", "1981-90", "1991-00",
    "2001-06"), c("Europe", "Asia", "Americas",
    "Africa", "Oceania")))
```
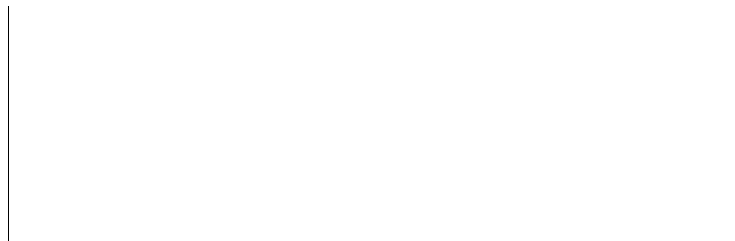
The regional data is presented in columns. We could use it in this format and manually add all the regional series, however with little data manipulation it is possible to considerably speed up the following steps.

```
> df.m <- melt(df)
> df.m <- rename(df.m, c(X1 = "Period", X2 = "Region"))
```

Now everything is set to start designing the plot.

```
> a <- ggplot(df.m, aes(x = Period, y = value/1e+06,
     fill = Region)) + opts(title = "Migration to the United States by Source Region (1820-2006)") +
     labs(x = NULL, y = "Number of People (in millions)n",
         fill = NULL)
> b <- a + geom_bar(stat = "identity", position = "stack")
```
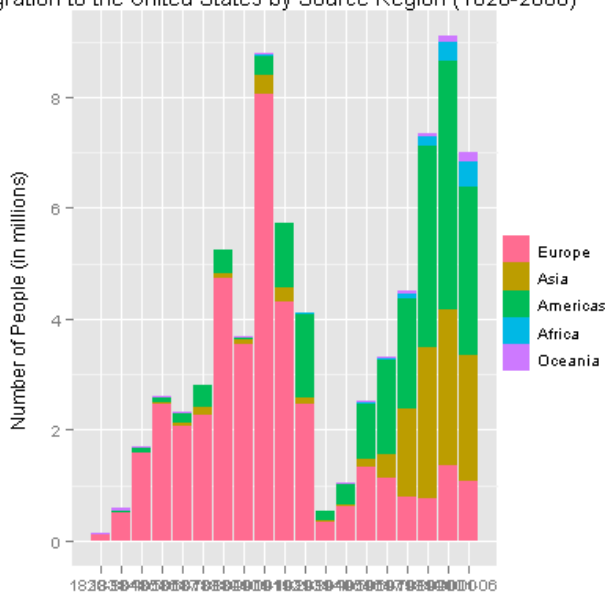

There are a few problems with this plot: the title is not completely visible, also x-axes labels are unreadable. We will fix them after first changing the colour palette.

The default colours are nice, still I prefer **ColorBrewer (http://www.personal.psu.edu/cab38/ColorBrewer/ColorBrewer_intro.html)** palettes.

```
> b <- b + scale_fill_brewer(palette = "Set1")
```

Migration to the United States by Source Region (1820-2006)
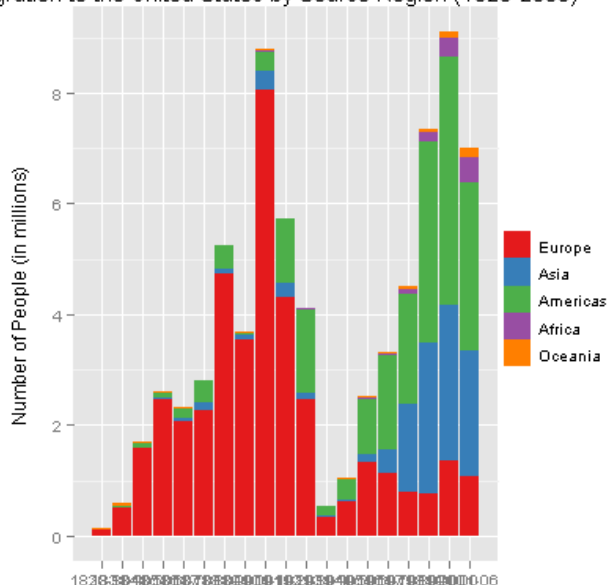
Next we will make changes to the default plotting theme used, this will also take care of the few problems highlighted above.

```
> immigration_theme <- theme_update(axis.text.x = theme_text(angle = 90,
    hjust = 1), panel.grid.major = theme_line(colour = "grey90"),
    panel.grid.minor = theme_blank(), panel.background = theme_blank(),
    axis.ticks = theme_blank(), legend.position = "none")
> b
```
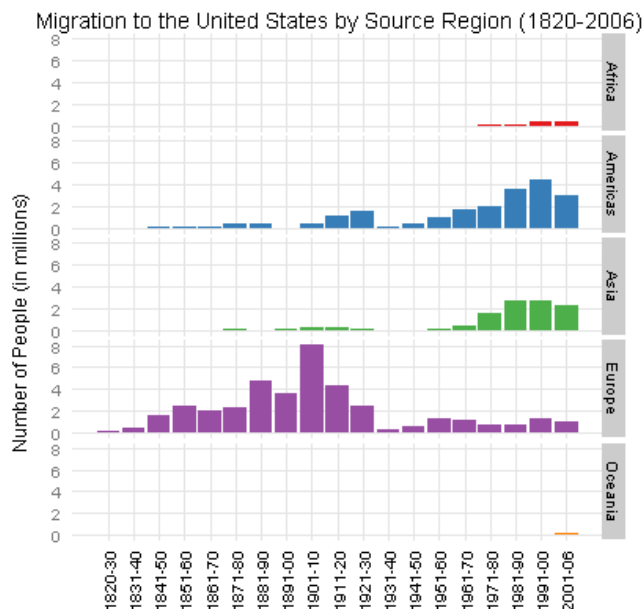

Migration to the United States by Source Region (1820-2006)
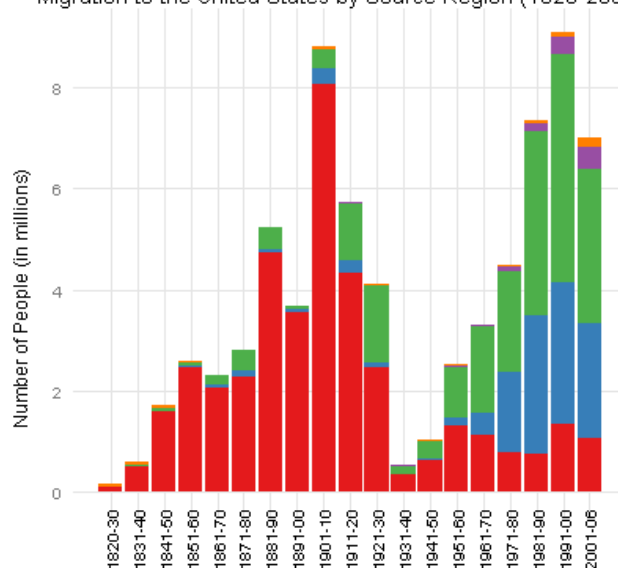
# Alternative visualisations

Next I would also like to present two alternative views of the same data using faceting capabilities of ggplot2.

```
> c <- b + facet_grid(Region ~ .) + opts(legend.position = "none")
```

Migration to the United States by Source Region (1820-2006)

Additionally, it would be interesting to see the change in total immigration on the same graph. In order to do this, a new dataframe with the annual totals will be created and later merged with the existing dataset (variable names in both dataframes should be identical for this to work). Then we just change the dataframe the plot is based on.

```
> total <- cast(df.m, Period ~ ., sum)
> total <- rename(total, c(`(all)` = "value"))
> total$Region <- "Total"
> df.m.t <- rbind(total, df.m)
> c1 <- c %+% df.m.t
```


Migration to the United States by Source Region (1820-2006)
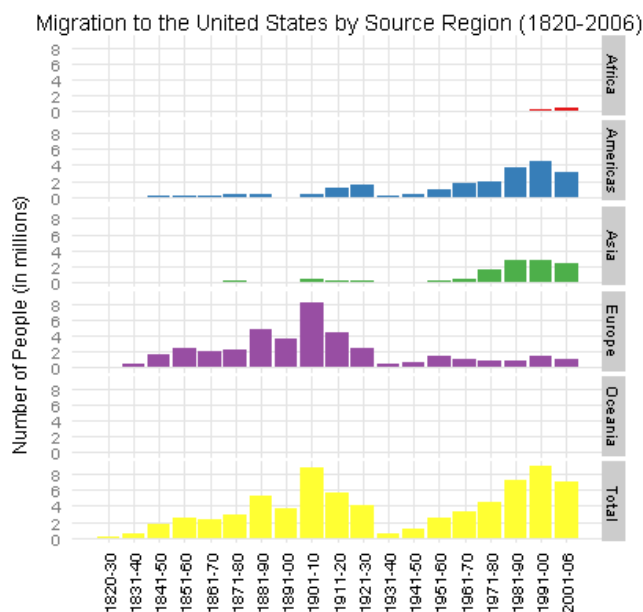
Update: Hadley suggested in the comments to "try scale="free_y" and possibly space="free" so it's possible to see the counts for the smaller regions."

Here goes:

```
> c2 <- c1 + facet_grid(Region ~ ., scale = "free_y")
```

Migration to the United States by Source Region (1820-2006)
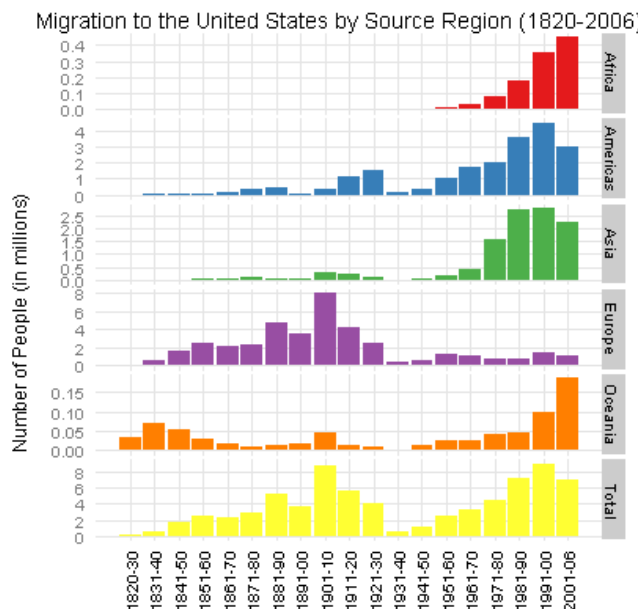
Use of opts()

Next, note how the formatting is changed with the opts() function instead of a theme update.

```
> d <- a + geom_bar(aes(x = Region, y = value/1e+06)) +
    theme_grey() + facet_wrap(~Period) + scale_fill_brewer(palette = "Set1") +
    opts(axis.text.x = theme_text(angle = 90,
        hjust = 1)) + opts(axis.ticks = theme_blank()) +
    opts(legend.position = "none") + opts(panel.grid.minor = theme_blank())
```

Migration to the United States by Source Region (1820-2006)
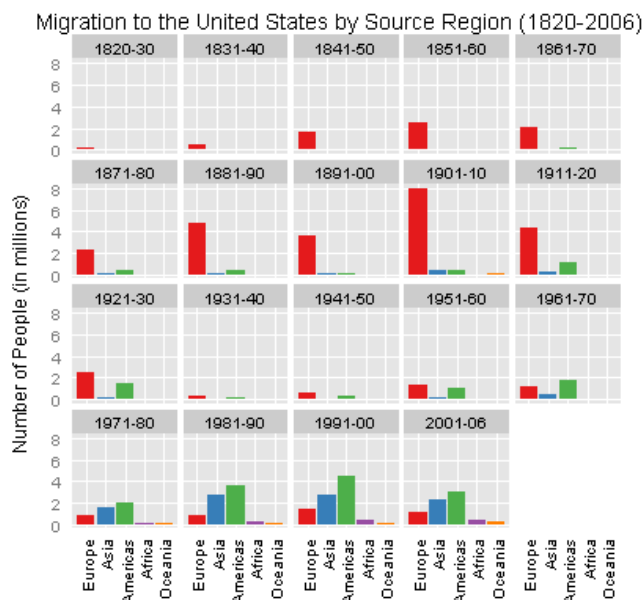
20 Comments   leave one →

1. **Hadley Wickham PERMALINK**
   March 16, 2009 7:12 pm

   Great idea for a blog

   In the second to last plot you might want to try scale="free_y" and possibly space="free" so it's possible to see the counts for the smaller regions.

   REPLY

2. **romunov PERMALINK**
   August 20, 2009 11:14 pm

It would seem I can't reproduce your work. Everything goes smooth until I try to draw the graph (b). I get a blank plot area. Any ideas?

REPLY

3. **romunov PERMALINK**
August 21, 2009 12:55 am
Update:

This is the error message I get when trying to execute "b".

```
Error in get("transform", env = ., inherits = TRUE)(., ...) : attempt to apply non-function
```

When trying to plot "a", I get:
```
Error: No layers in plot
```
Which is the same error I get when I try to plot my own data.

REPLY

o **learnr PERMALINK***
August 22, 2009 6:13 pm
The error message for "a" is correct – there really are no layers to plot – these are only added in plotting "b" (geom_bar).

As to the other error message: check that you have RColorBrewer installed?

REPLY

4. **romunov PERMALINK**
August 22, 2009 7:37 pm
I've installed and loaded RColorBrewer into R but I still get the same error message. I run a version of 2.9.1 on WinXP. A friend of mine tried it in Ubuntu (9.04) and it works fine (out of the box).

REPLY

o **learnr PERMALINK***
August 23, 2009 11:02 am
I am also running R 2.9.1 on XP.

Do the examples on http://had.co.nz/ggplot2/geom_bar.html work? These examples are using built-in datasets.

If they do work, then there is a problem importing your data. More often than not the problem is the formatting of quotation marks. Check that `str(df)` is meaningful.

REPLY

5. **romunov PERMALINK**
August 23, 2009 11:13 am
I suspect this is ggplot/R related and does not pertain my data. I've tried the ggplot samples using internal data and it yields an identical error:

```
> c c + geom_bar()
Error in get("transform", env = ., inherits = TRUE)(., ...) :
attempt to apply non-function
```

I really appreciate you taking the time with this.

REPLY

o **learnr PERMALINK***
August 23, 2009 11:57 am
Just in case – does the following also give you an error message?
```
ggplot(mtcars, aes(factor(cyl))) + geom_bar()
```

Try reinstalling both R and ggplot2 – it seems that something is broken.

REPLY

6. **romunov PERMALINK**
August 23, 2009 11:52 am
Additionally, I'm getting some funky errors as well which is leading me to believe that this is not my data problem, but something else. Perhaps R, ggplot2 or something else.

```
> d <- ggplot(diamonds, aes(carat, price, fill = ..density..)) +
+ + xlim(0, 2) + stat_binhex(na.rm = TRUE) + opts(aspect.ratio = 1)
Error in +xlim(0, 2) : invalid argument to unary operator
```

REPLY

7. **romunov** **PERMALINK**
August 23, 2009 12:03 pm
Hehe, great minds think alike. As you've posted the code for plotting mtcars I've also sent a message to ggplot2 googlegroup that this isn't working.

```
> ggplot(mtcars, aes(factor(cyl))) + geom_bar()
Error in get("transform", env = ., inherits = TRUE)(., ...) :
attempt to apply non-function
```

I will try reinstalling R and ggplot. Upgrade from 2.8.1 two days ago didn't do the trick, I guess.

REPLY

8. **romunov** **PERMALINK**
August 23, 2009 1:30 pm
No go after a fresh install.

REPLY

9. **romunov** **PERMALINK**
October 2, 2009 9:00 pm
I'm happy to inform you that upgrading R to 2.9.2 solved my ggplot2 debacle. Hopefully I will be plotting ggplot2 graphs in no time.

REPLY

10. **JohnMajor** **PERMALINK**
November 11, 2009 3:41 am
Exactly the example I was looking for! Thanks!

REPLY

11. **Attachai Jintrawet** **PERMALINK**
April 9, 2011 6:44 am
How to plot bar graph (for example daily rainfall) with scatter plot (for example daily air temperature), using Julian date as x-axis.

Thank you,
Attachai

REPLY

- **learnr** **PERMALINK*
April 12, 2011 4:27 pm
Sorry, I struggle to understand what you are trying to do. Could you point me to an example on the web? Or are you referring to facet plots?

REPLY

12. **Denis Chabot** **PERMALINK**
April 17, 2011 4:10 pm
Hi, Thanks for the example and how to produce it. However I was unable to get the first "b" to work. I got the error message "Erreur dans titles[[i]] : indice hors limites" (something like "Error in titles[[i]]: off-limit indices") and no plot. However, the theme modification that follows makes "b" (this time with white instead of gray background) work correctly.

I asked why on the ggplot2 mailing list and Kohske Takahashi suggested I replaced labs(fill=NULL) by labs(fill=""). His suggestion worked. Since the theme modification removes the legend all together, the problem disappeared after it.

I do not know why the example worked as is in 2009 and not now. Maybe the difference is due to changes in ggplot2 or in R. I use ggplot2 0.8.9 and R 2.13 on Mac OS X.

Denis

REPLY

- **learnr** **PERMALINK*
April 18, 2011 9:37 pm
Thanks, you're right the syntax must have changed since then, and `labs(fill="")` really works and `labs(fill=NULL)` doesn't.

REPLY

13. **stolzyblog** **PERMALINK**
June 29, 2013 8:22 pm
Nice article! I like the various multiple plot per page examples. With barplots I often catch myself thinking about ways to display other dimensions of info (provided the data were available). For example, what are some ways we could depict target areas of the U.S. migrated into over the decades (northeast; midwest, south, great lakes, west, etc.) simultaneously with the other currently

displayed info for source region, quantity, and year? As another example, I had a problem I was trying to solve involving displaying multiple 'scales' on the y-axis recently. I came up with the ggplot2 solution here, using ANNOTATE, but maybe someone sees a different or less wordy way: http://tinyurl.com/q2dvks8

~RS

REPLY

Trackbacks

1. Blog
2. Layering in R for Oracle Data « Hearing the Oracle

Create a free website or blog at WordPress.com.

The Vigilance Theme.

Follow

Follow "Learning R"

Powered by WordPress.com