

Bootcamp 6: Area plots

Lucien Baumgartner

10/30/2018

NOTE: Old script from last year.

The important part is the `expand.grid()` function.

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
library(scales)

setwd("~/ddj18/output/")

rm(list=ls())

load('01-bevoelkerung-clean.RData')

jx <- df %>%
  filter(!alterv05kurz %in% c('0-4', '5-9', '10-14', '15-19')) %>% # U18-jährige rausfi
  group_by(stichtagdatjahr, geblandhistlang, ziv2lang) %>% # wir gruppieren dan den Aggr
  summarise(n_grp=n()) %>% # Berechnung der ANZAHL Personen mit Zivstatus X in Land X i
  mutate(n_total=sum(n_grp), # Total berechnen; 100% = Alle Personen aus Land X im Jahr
         perc=n_grp/n_total) %>% # Anteile berechnen
  ungroup
jx
```

```
## # A tibble: 2,623 x 6
##   stichtagdatjahr geblandhistlang ziv2lang   n_grp n_total   perc
##         <int> <chr>           <chr>     <int>   <int>   <dbl>
## 1      1993 Afrika           Geschieden    395    2827 0.140
## 2      1993 Afrika           Ledig         641    2827 0.227
## 3      1993 Afrika       Verheiratet   1740    2827 0.615
## 4      1993 Afrika       Verwitwet      51    2827 0.0180
## 5      1993 Amerika       Geschieden    350    2998 0.117
## 6      1993 Amerika           Ledig        771    2998 0.257
## 7      1993 Amerika       Verheiratet  1817    2998 0.606
## 8      1993 Amerika       Verwitwet     60    2998 0.0200
## 9      1993 Asien         Geschieden    473    5582 0.0847
## 10     1993 Asien           Ledig       1402    5582 0.251
## # ... with 2,613 more rows
```

```
# gewisse Länder haben in gewissen Jahren 0 Beobachtungen für Zivstatus X
# Daher bilden wir ein fiktives Datenset, das alle möglichen Wertkombinationen der Aggr
jgrid <- expand.grid(stichtagdatjahr=unique(jx$stichtagdatjahr),
                    geblandhistlang=unique(jx$geblandhistlang),
                    ziv2lang=unique(jx$ziv2lang))
# danach 'mergen' wir das Set mit allen möglichen Kombinationen mit den tatsächlichen W
# Wieso? Alle Kombinationen die im echten Datenset nicht vorkommen werden NA's bekommen
# Anzahl, Total und Prozente
jx <- full_join(jgrid, jx)
```

```
## Joining, by = c("stichtagdatjahr", "geblandhistlang", "ziv2lang")
```

```
## Warning: Column `geblandhistlang` joining factor and character vector,
## coercing into character vector
```

```
## Warning: Column `ziv2lang` joining factor and character vector, coercing
## into character vector
```

```

# diese NA's ersetzen wir durch Nullen -> das brauchen wir für die Area Plots, da sonst
jx <- mutate_at(jx, c('n_grp', 'n_total', 'perc'), function(x){ifelse(is.na(x), 0, x)})

# Disclaimer: Diese Schritte sind eigentlich nicht mehr nötig, das wir eh nur Länder be
# jährlich mehr als 400 Beobachtungen haben; habs trotzdem drin gelassen

## Plot 1
# Wir machen einen Vektor, der die Herkunftsländer nach der grössten Summe von Geschied
# brauchen wir, um die Facets im Plot danach zu ordnen
lvls <- jx %>% filter(ziv2lang=='Geschiedene') %>% # nur Geschiedene
  # alle diese Länder wollen wir nicht:
  filter(!geblandhistlang %in% c("Bosnien und Herzegowina",
                                "Mazedonien",
                                "Serbien",
                                "Serbien, Montenegro, Kosovo",
                                "Afrika",
                                "Asien",
                                "Amerika",
                                "Niederlande",
                                "Österreich",
                                "Ozeanien",
                                "Unzuteilbar",
                                "Kroatien",
                                "Italien",
                                "Europa"),
         !is.na(ziv2lang)) %>% #keine NA's!
  group_by(geblandhistlang, ziv2lang) %>% # aggregieren
  summarise(x=sum(perc)) %>% # Summe der Anteile geschiedener Personen berechnen
  arrange(desc(x)) %>% # absteigend anordnen
  select(geblandhistlang) %>% # nur Länder isolieren
  unlist # zu Vektor transformieren

q <- jx %>%
  # die folgenden 2 Levels wollen wir nicht in der Zivstand Variable
  mutate(ziv2lang=ifelse(ziv2lang=="Aufgelöste Partnerschaft", NA, ziv2lang),
         ziv2lang=ifelse(ziv2lang=="Unbekannt", NA, ziv2lang)) %>%
  # diese Länder uch nicht
  filter(!geblandhistlang %in% c("Bosnien und Herzegowina",
                                "Mazedonien",
                                "Serbien",
                                "Serbien, Montenegro, Kosovo",
                                "Afrika",
                                "Asien",
                                "Amerika",
                                "Niederlande",
                                "Österreich",
                                "Ozeanien",
                                "Unzuteilbar",
                                "Kroatien",
                                "Italien",
                                "Europa"),
         !is.na(ziv2lang)) %>% # NA's auch nichz
  mutate(geblandhistlang=factor(geblandhistlang, levels=lvls)) %>% # wir machen Zivsta

# Plot:
ggplot(., aes(x=stichtagdatjahr, y=perc, fill=ziv2lang, group=ziv2lang)) +

```

```

geom_area(position = "stack") +
scale_fill_manual(values=c('firebrick1','dodgerblue', 'dodgerblue4', 'gold1', 'darkgray',
                           name='Zivilstand')) +
labs(caption='(XXX; Politischer Datenjournalismus XXX)',
      title="Immer weniger Verheiratete in der Stadt Zürich",
      subtitle="Anteil Personen mit Zivilstand x über die Zeit, nach Geburtsland") +
scale_x_continuous(breaks=c(1993, 2000, 2010, 2016)) +
scale_y_continuous(labels=percent) +
facet_wrap(~geblandhistlang, nrow = 5, ncol = 3) +
theme(panel.background = element_blank(),
      text = element_text(size=15),
      plot.title = element_text(face='bold')) +
ylab('Anteil') +
xlab('Jahr')
q # plot printen

```

