

Interaktives Sankey-Diagramm mit R

Benjamin Schlegel

31. Oktober 2016

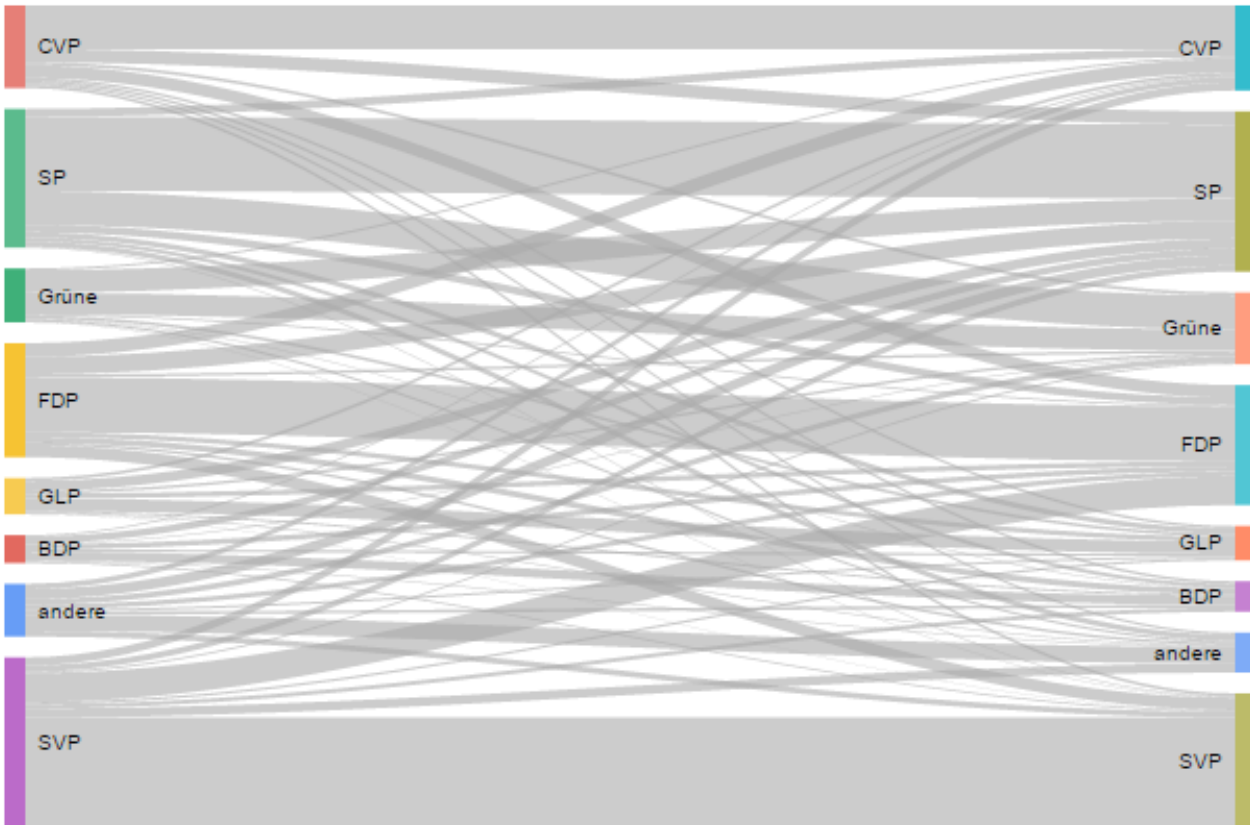
In R gibt es mehrere Möglichkeiten ein Sankey-Diagramm zu erstellen. Unter anderem mit [rCharts](#) und [googleVis](#). In diesem Artikel wird ein Beispiel mit [googleVis](#) gezeigt.

Als erstes muss der Datensatz eingelesen werden. Der Datensatz "National- vs. Ständeratswahlen 2015" kann unter [Data](#) heruntergeladen werden. Anschliessend kann er mit [read.csv](#) eingelesen werden. Nach dem einlesen werden noch die Variablennamen geändert und bei der Variable sr ein Leerschlag hinzugefügt, damit es zu keiner Schlaufe beim Sankey-Diagramm kommt.

```
data = read.csv("nr_sr15.csv")
colnames(data) = c("nr", "sr", "weight", "tooltip")
data$sr = paste0(data$sr, " ")
```

Nun sind die Daten bereit. Um ein Sankey-Diagramm zu zeichnen, muss zuerst die Bibliothek [googleVis](#) geladen werden. Anschliessend kann das Diagramm mit der Funktion [gvisSankey](#) gezeichnet werden. Mit [plot](#) um die Funktion herum, wird direkt eine HTML-Datei im Browser angezeigt und das Kunstwerk kann betrachtet werden. `data[,1:3]` wird verwendet, da nur die ersten drei Variablen aus dem Datensatz gebraucht werden.

```
require(googleVis)
plot(gvisSankey(data[,1:3], from="NR", to="SR", weight="Freq"))
```



Data: data[, 1:3] • Chart ID: SankeyID1af81f3f1c1e • googleVis-0.6.1
 R version 3.3.1 (2016-06-21) • [Google Terms of Use](#) • [Documentation and Data Policy](#)

Als nächstes werden Farben definiert, damit die Parteifarben auch korrekt sind. Dazu wird ein Vektor mit den Farben definiert. Die Reihenfolge muss ausprobiert werden, die sie keine Logik hat (resp. der Autor sie noch nicht gefunden hat). Mit dem restlichen Code werden die Farben in die JSON Form gebracht und dem Diagramm hinzugefügt.

```
colors_node =
c('grey', 'grey', '#000000', '#FF850C', '#009FDA', '#B4DC00', '#FF448A', '#ED1C24', '#148A3C',
  '#000000', '#FF850C', '#009FDA', '#B4DC00', '#FF448A', '#ED1C24', '#148A3C')
colors_node_array = paste0("[", paste0("'", colors_node, "'", collapse = ', '), "]")
opts = paste0("{
  link: { colorMode: 'target' },
  node: { colors: ", colors_node_array, " }
}")

plot(gvisSankey(data[,1:3], from="NR", to="SR", weight="Freq",
  options = list(sankey=opts)))
```

Nun sieht das Sankey-Diagramm schon recht ansprechend aus. Was jedoch noch stört sind die unschönen Labels. Auch das kann angepasst werden. Dazu muss der ganze Datensatz verwendet werden.

```
plot(gvisSankey(data, from="NR", to="SR", weight="Freq",
  options = list(sankey=opts)))
```

Jetzt wird eine Fehlermeldung angezeigt. Das liegt daran, dass google nichts mit den Labels anfangen kann. Um das zu ändern, muss die HTML gespeichert werden und anschliessend in einem Editor (z.B. Notepad++, Windows Editor) geöffnet werden. Dort muss die Zeile 419 `data.addColumn('string','tooltip');` durch den Code `data.addColumn({type: 'string', role: 'tooltip'});` ersetzt werden. Anschliessend kann die

HTML-Datei gespeichert werden und in einem Browser geöffnet werden.

Nun können noch die Zeilen 492 bis 497 gelöscht werden, um den Text unterhalb des Diagramm zu entfernen. Nun kann die HTML-Datei mit [iframe](#) in einem Blog eingebettet werden.<div align="center" style=" position: relative; height: 0; overflow: hidden; width: 100%; height: 600px">