Basic Graphs for Variables in Datasets

Most of the graphs for datasets use the formula interface[1] and are part of the lattice package, which is loaded together with the hanoverbase package[2].

- We will be using the ggformula package, which works well the ggplot package while providing a basic formula interface to the different graphs.
- It is hard to produce pie and pareto charts this way. You can look at the basic cheatsheet[3] for ways to produce those kinds of charts.
- ggformula commands expect the dataset as their first argument, and therefore can work well with the pipe operator %>%[4]

Use `vignette("ggformula")` within RStudio or this link[5] to see more examples of ggformula.

Graphs for categorical variables

Common graphs for categorical variables are dotplots and barcharts.

```
library(hanoverbase)
library(ggformula)
data(brfss)

gf_counts(brfss, ~genhealth)
brfss %>% gf_bar(~genhealth)
brfss %>% gf_barh(~genhealth)
brfss %>% gf_countsh(~genhealth, fill="blue", col="black")
brfss %>% gf_countsh(~genhealth, geom="point")
brfss %>% gf_percentsh(~genhealth)
brfss %>% gf_percents(~genhealth)
```

Graphs for quantitative variables

Common graphs for quantitative variables are histograms and box-and-whisker plots.

```
library(hanoverbase)
library(ggformula)
data(counties)

counties %>% gf_histogram(~female)
counties %>% gf_boxplot(~female)
counties %>% gf_boxploth(1~female)
```

[1] formulas.html
[2] packagesAsToolboxes.html
[3] ../cheatSheet.html
[4] piping.html
[5] https://cran.r-project.org/web/packages/ggformula/vignettes/ggformula.html