Ggplot2 tutorial - Command lines Part 3

Ariane Ducellier

10/17/2023

Load R packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
#library(gridExtra)
#library(Lock5Data)
#library(maps)
#library(mapproj)
```

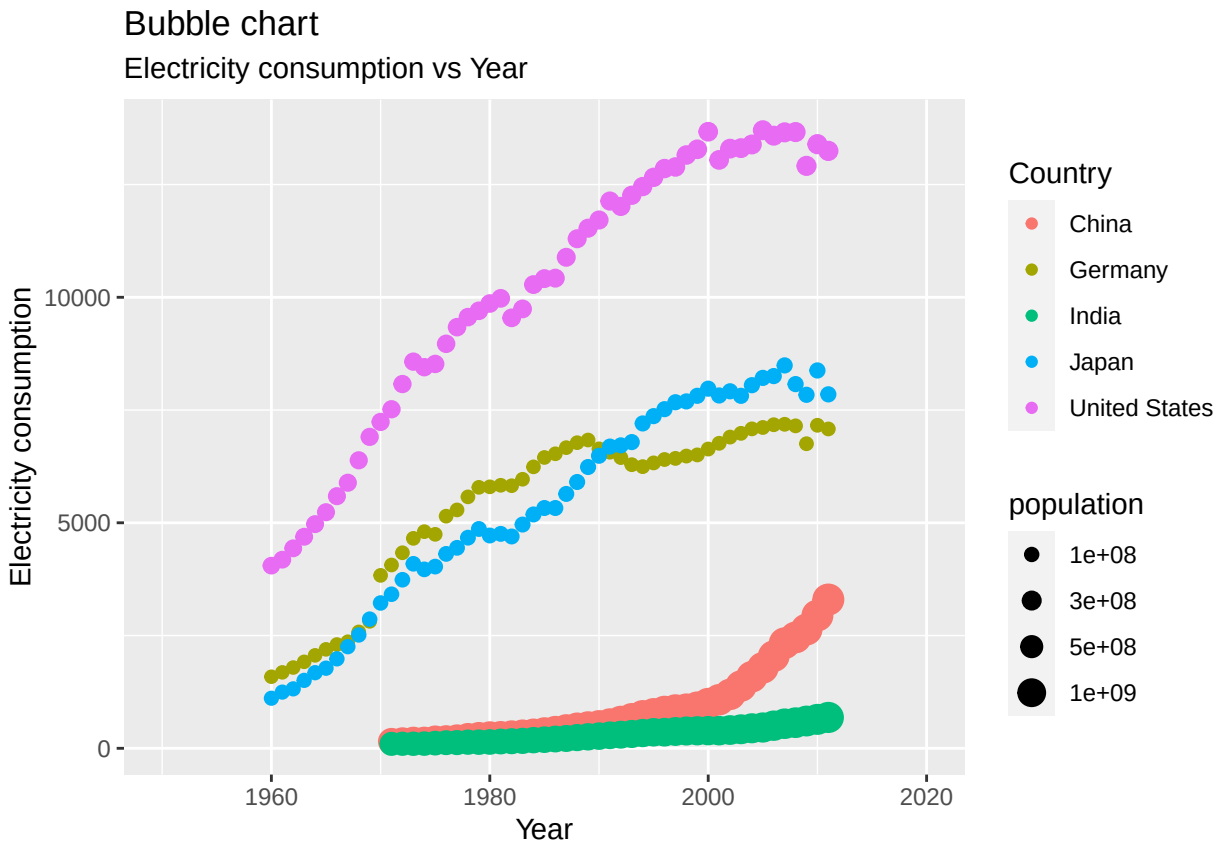
Part 3 - Advanced Geoms and Statistics

Bubble charts

```
df <- read.csv("../data/gapminder-data.csv")
dfs <- subset(df, Country %in% c("Germany", "India", "China", "United States", "Japan"))

ggplot(dfs, aes(x=Year, y=Electricity_consumption_per_capita)) + geom_point(aes(size=population, color=C
coord_cartesian(xlim=c(1950, 2020)) +
labs(subtitle="Electricity consumption vs Year", title="Bubble chart") +
ylab("Electricity consumption") +
scale_size(breaks=c(0, 1e+8, 0.3e+9, 0.5e+9, 1e+9, 1.5e+9), range=c(1, 5))
```

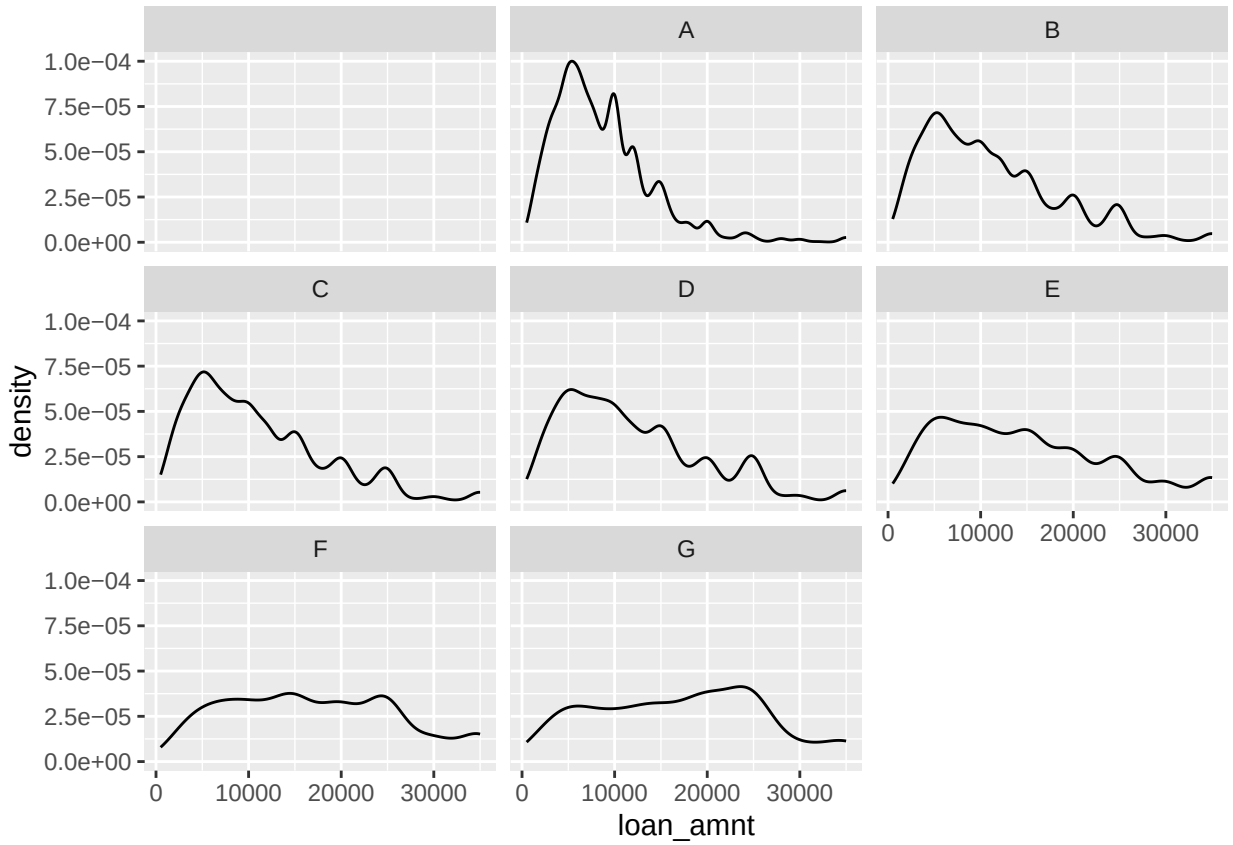
```
## Warning: Removed 842 rows containing missing values (`geom_point()`).
```



Density plots

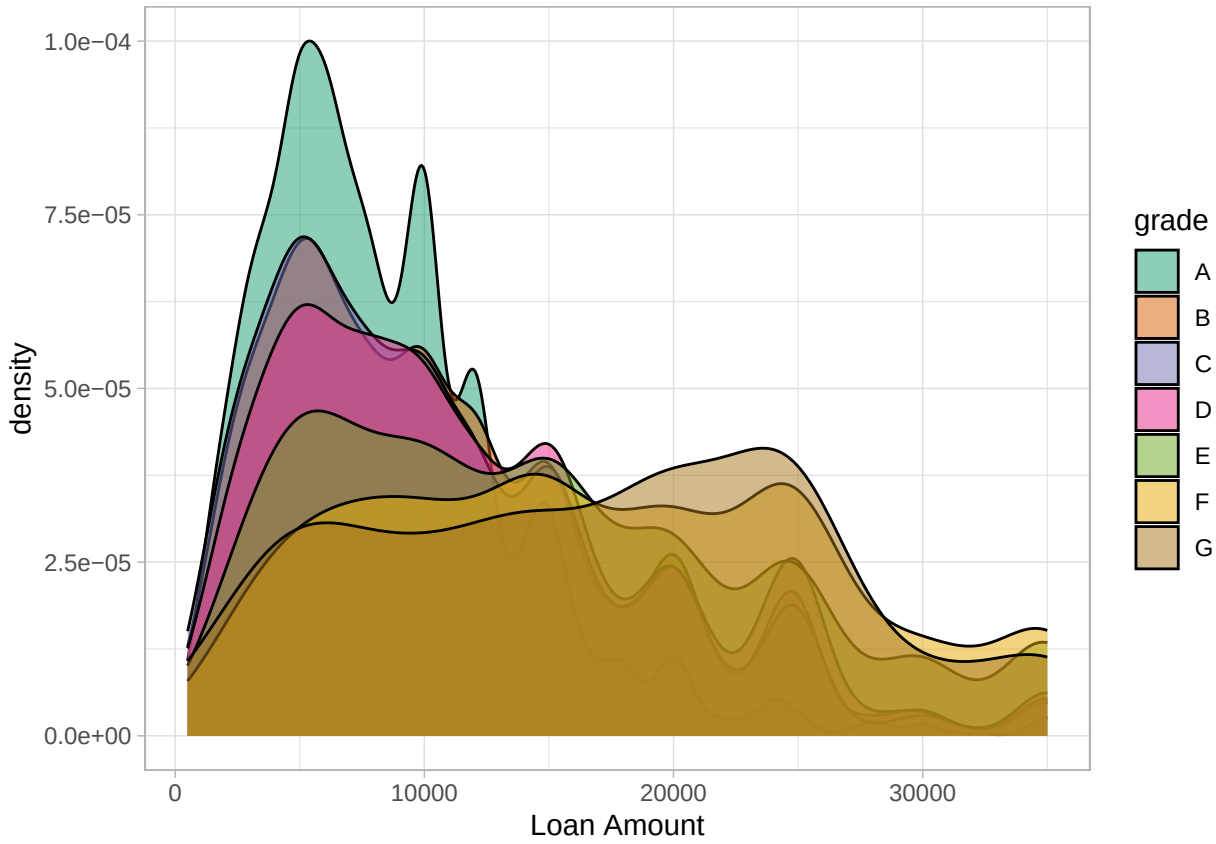
```
df <- read.csv("../data/LoanStats.csv")
ggplot(df, aes(x=loan_amnt)) + geom_density() + facet_wrap(~grade)
```

```
## Warning: Removed 7 rows containing non-finite values (`stat_density()`).
```



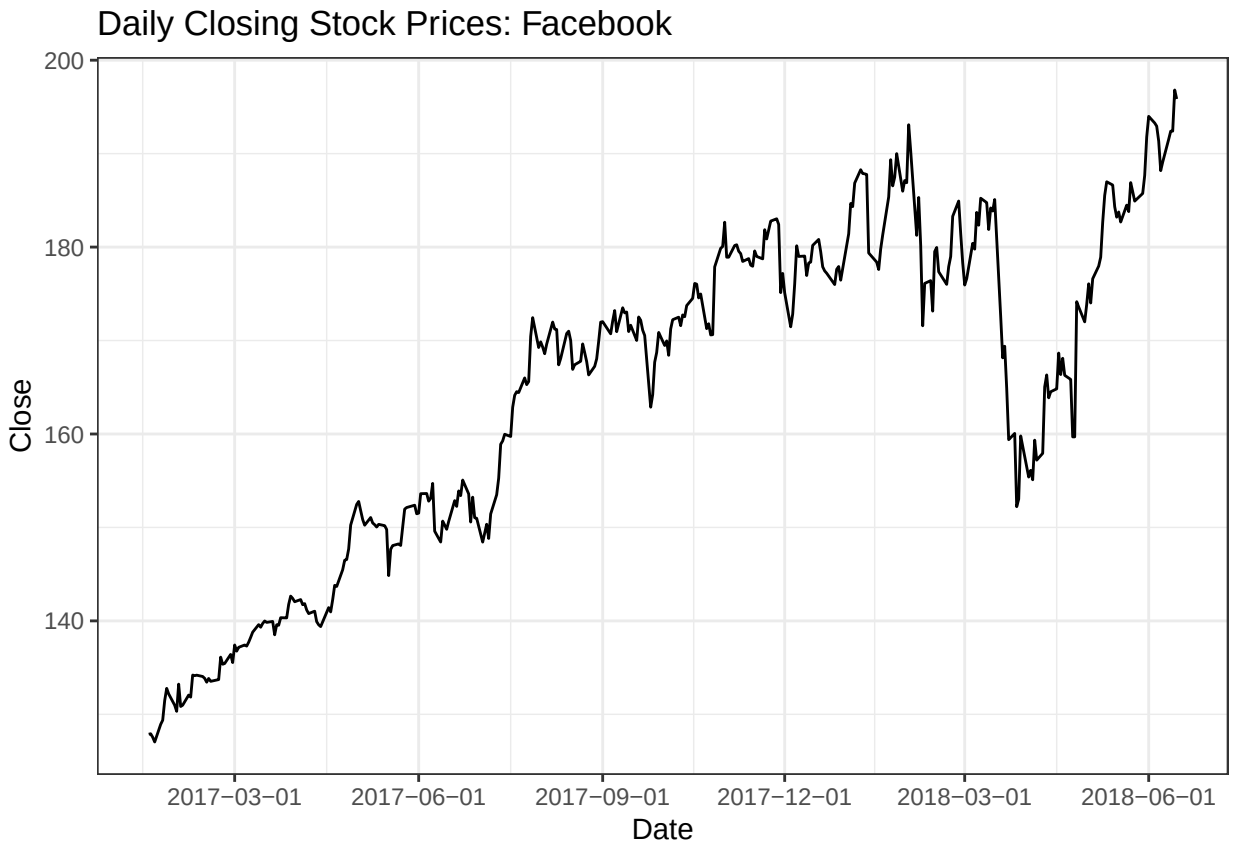

```
df <- read.csv("../data/LoanStats.csv")
ggplot(df, aes(x=loan_amnt)) + geom_density(aes(fill=grade), alpha=1/2) +
scale_fill_brewer(palette="Dark2") + xlab("Loan Amount") + theme_light()
```

```
## Warning: Removed 7 rows containing non-finite values (`stat_density()`).
```



Time series

```
df_fb <- read.csv("../data/FB.csv")
df_fb$Date <- as.Date(df_fb$Date)
ggplot(df_fb, aes(x=Date, y=Close, group=1)) +
geom_line(color="black", na.rm=TRUE) +
ggtitle("Daily Closing Stock Prices: Facebook") +
theme(plot.title = element_text(lineheight=.7, face="bold")) +
scale_x_date(date_breaks='3 month') +
theme_bw()
```



Statistical summaries

```
df_fb <- read.csv("../data/FB.csv")
df_fb$Date <- as.Date(df_fb$Date)
df_fb$Month <- strftime(df_fb$Date,"%m")
df_fb$Month <- as.numeric(df_fb$Month)
ggplot(df_fb, aes(Month, Close)) +
geom_point(color="red", alpha=1/2, position=position_jitter(h=0.0, w=0.0)) +
stat_summary(geom="line", fun="mean", color="blue", size=1) +
stat_summary(geom="line", fun="median", color="orange", size=1) +
stat_summary(geom="line", fun="quantile", fun.args=list(probs=.1), linetype=2, color="green", size=1) +
stat_summary(geom="line", fun="quantile", fun.args=list(probs=.9), linetype=2, color="green", size=1) +
scale_x_continuous(breaks=seq(0, 13, 1)) +
ggtitle("Monthly Closing Stock Prices: Facebook") +
theme_classic()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

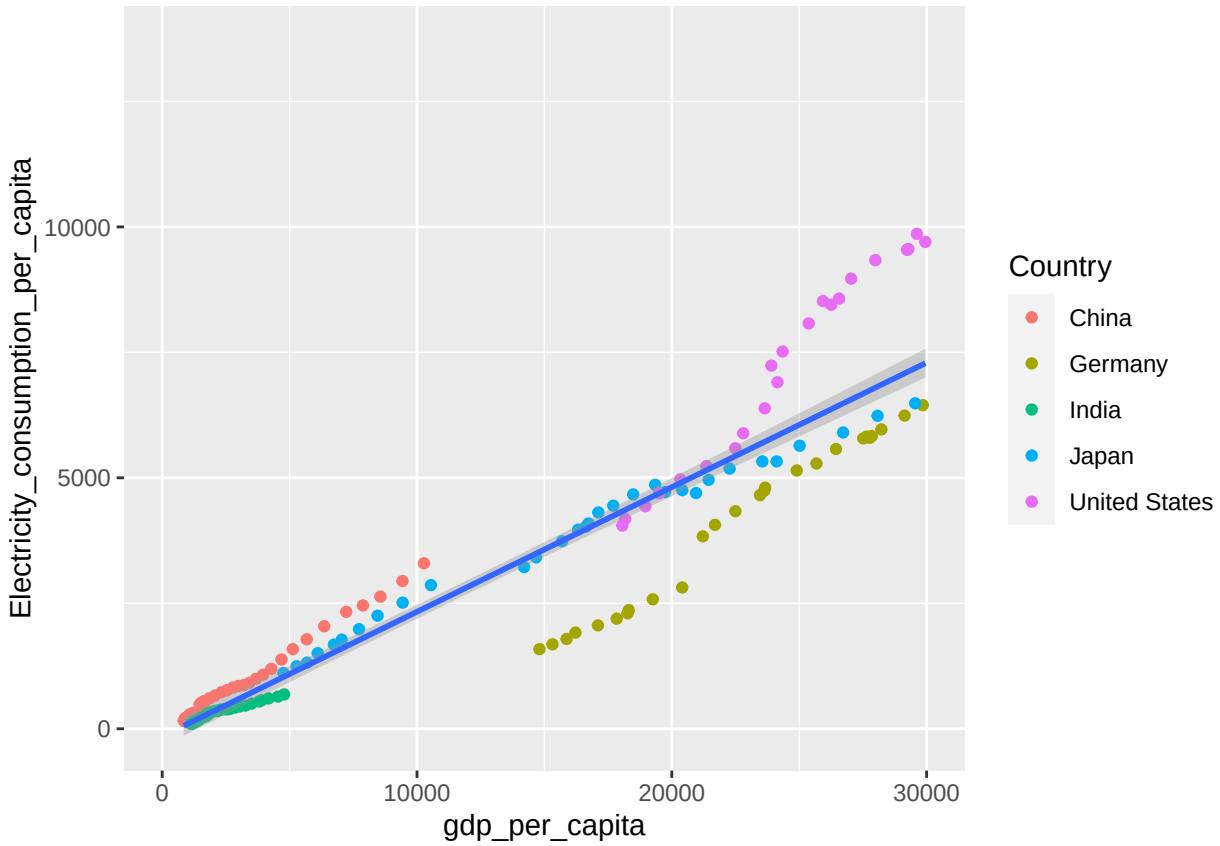
Linear regression

```
df <- read.csv("../data/gapminder-data.csv")
dfs <- subset(df, Country %in% c("Germany", "India", "China", "United States", "Japan"))
ggplot(dfs, aes(gdp_per_capita, Electricity_consumption_per_capita)) + geom_point(aes(color=Country)) +
xlim(0, 30000) +
stat_smooth(method=lm)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 919 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 919 rows containing missing values (`geom_point()`).
```

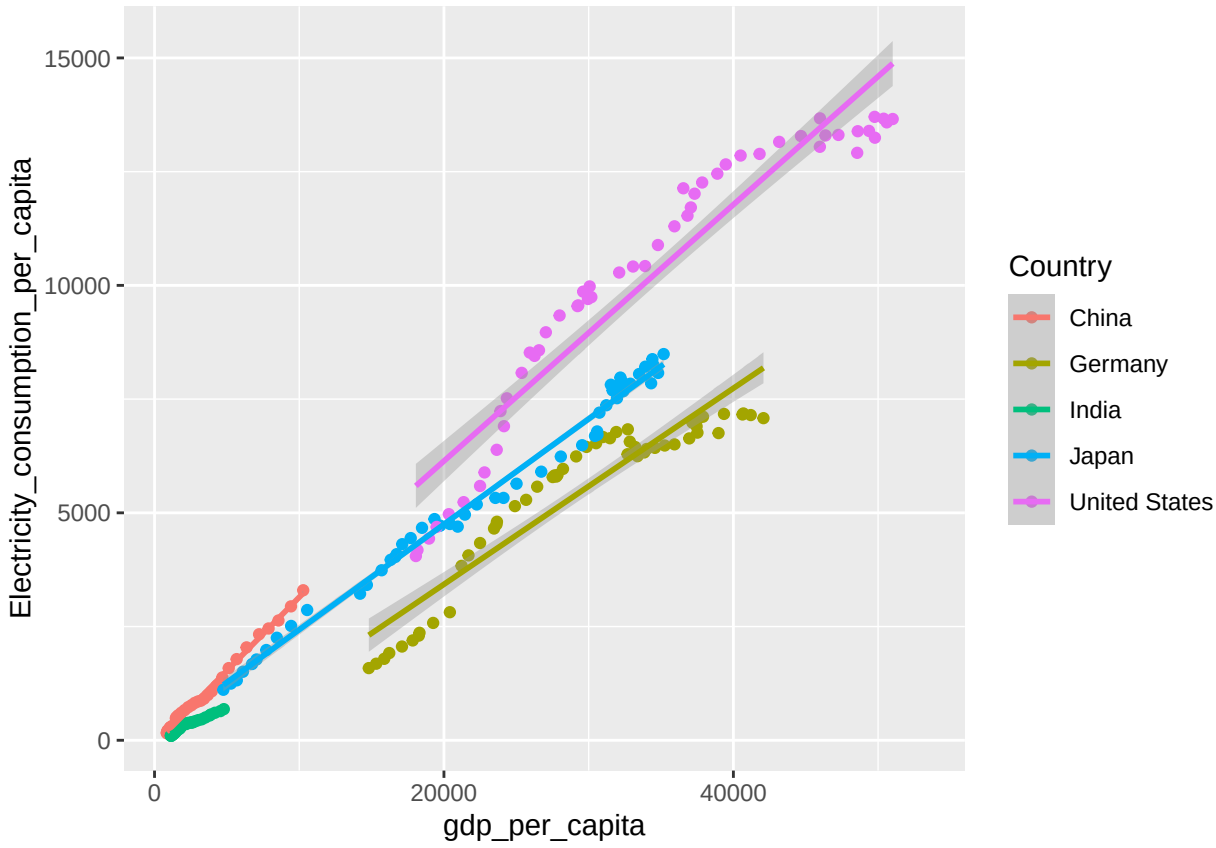



```
df <- read.csv("../data/gapminder-data.csv")
dfs <- subset(df, Country %in% c("Germany", "India", "China", "United States", "Japan"))
ggplot(dfs, aes(gdp_per_capita, Electricity_consumption_per_capita, color=Country)) +
geom_point() +
stat_smooth(method=lm)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

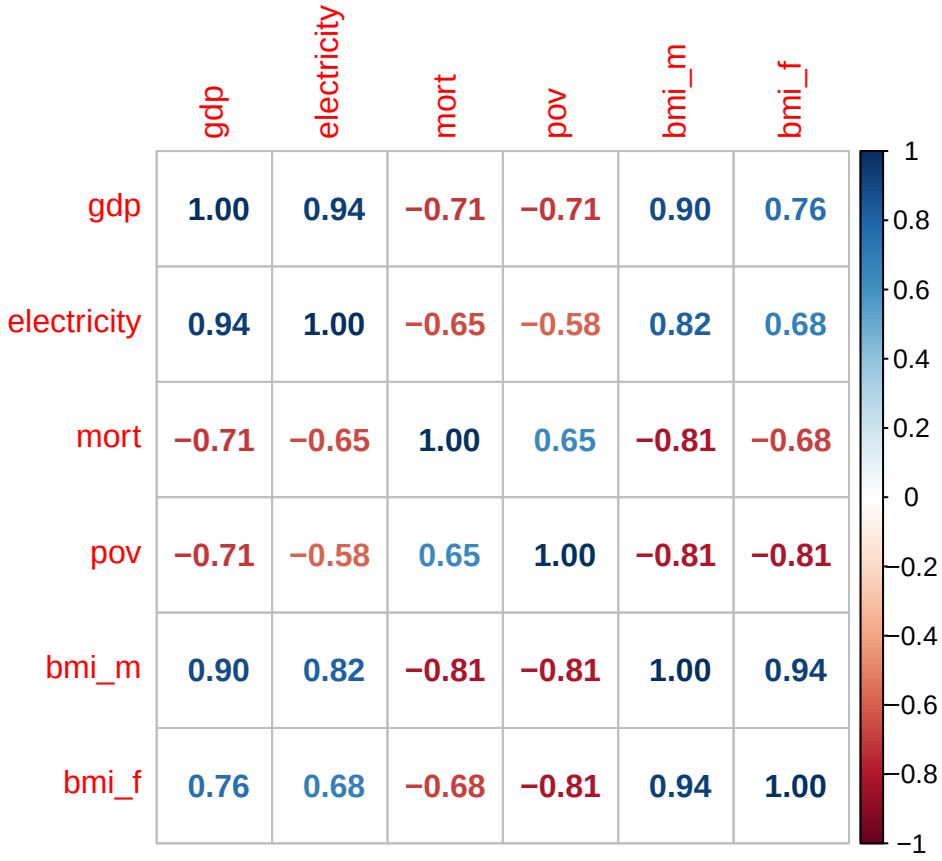
```
## Warning: Removed 842 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 842 rows containing missing values (`geom_point()`).
```

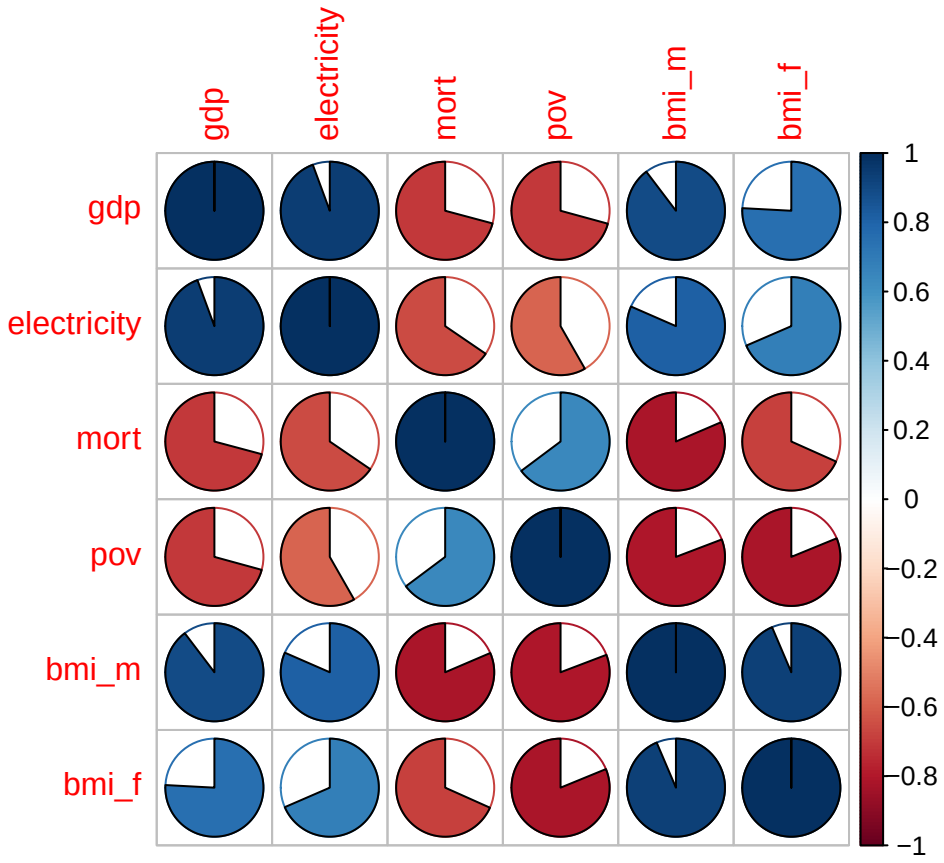


Correlations

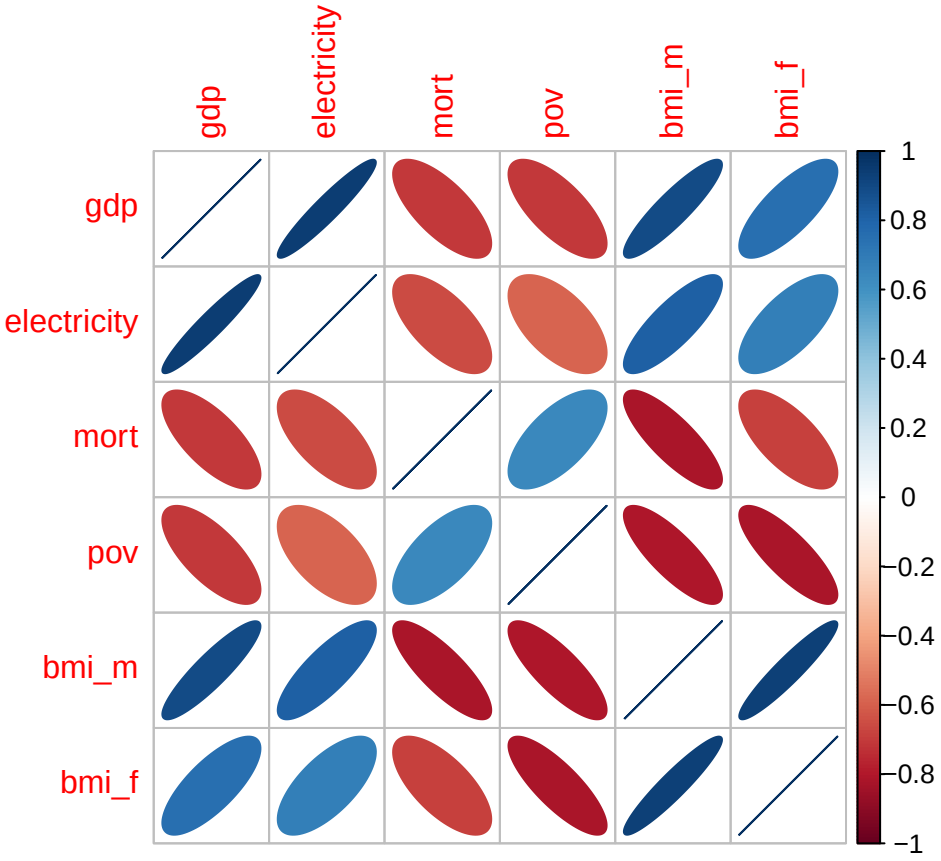
```
df <- read.csv("../data/gapminder-data.csv")
df <- df[, colnames(df)[4:9]]
df <- na.omit(df)
colnames(df) <- c("gdp", "electricity", "mort", "pov", "bmi_m", "bmi_f")
M <- cor(df)
corrplot(M, method="number")
```




```
corrplot(M, method="pie")
```

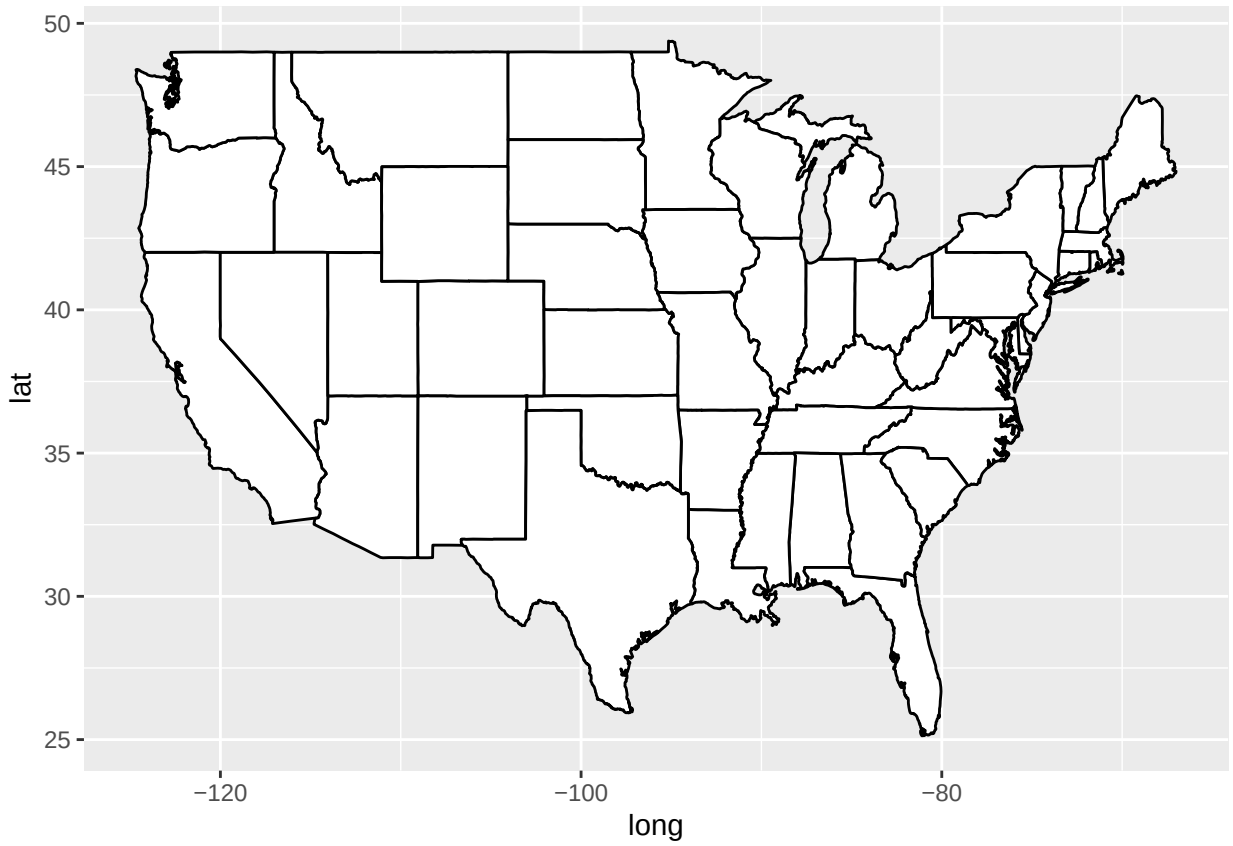



```
corrplot(M, method="ellipse")
```

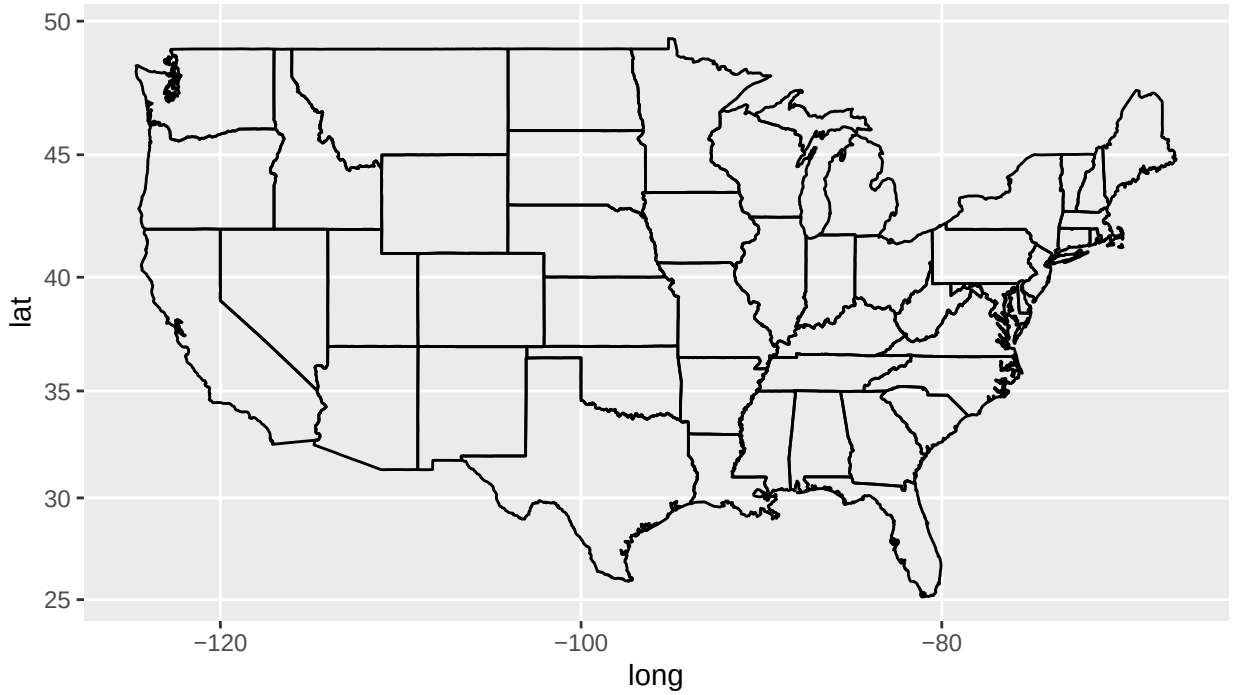


Maps

```
states_map <- map_data("state")
ggplot(states_map, aes(x=long, y=lat, group=group)) +    geom_polygon(fill="white", colour="black")
```




```
ggplot(states_map, aes(x=long, y=lat, group=group)) +
geom_path() + coord_map("mercator")
```




```
europe <- map_data("world",
    region=c("Germany", "Spain", "Italy", "France", "UK", "Ireland"))
```

```
ggplot(europe, aes(x=long, y=lat, group=group, fill=region)) +    geom_polygon(color="black") +
scale_fill_brewer(palette="Set3")
```