

Anhang 4. Projekt DLBDBSC01

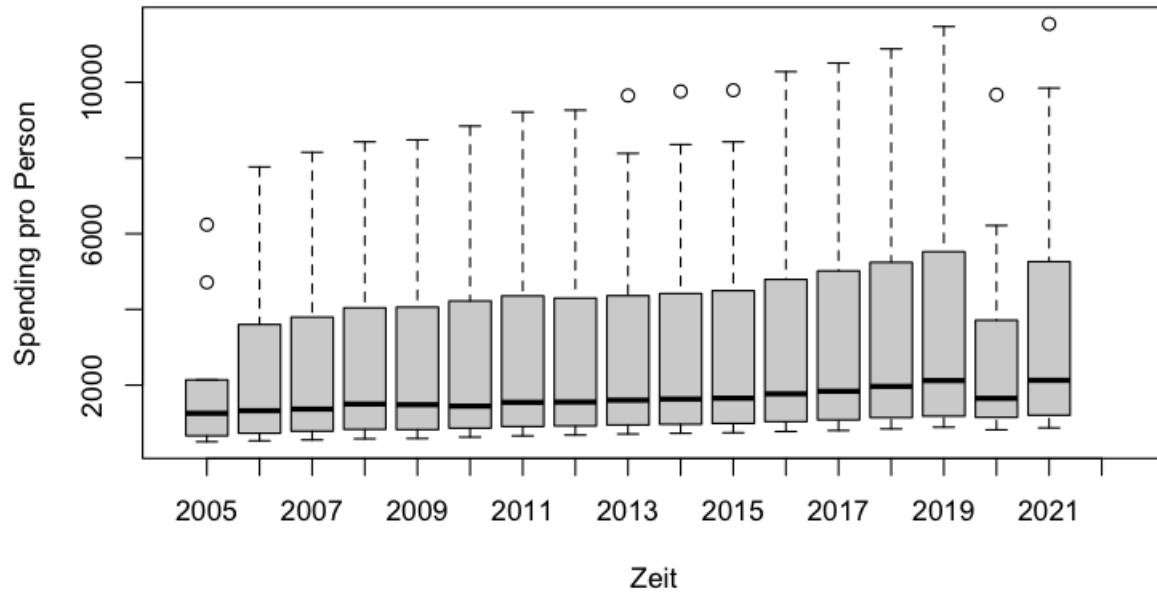
Code ▾

2023-04-12

Regressionsanalyse Enviroment: RData02 Lineare Regression für unabhängige Variable "Zeit" und abhängige Variable "SpendingProPerson" für top-10 Länder berechnen, Darstellung der Werte in Plot

Hide

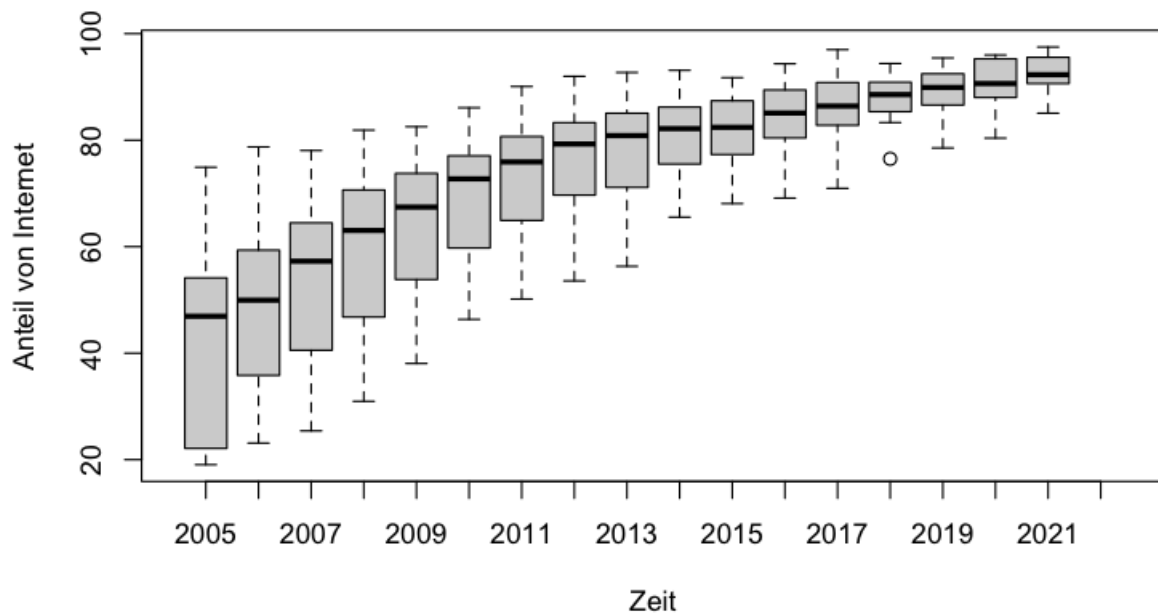
```
lm.spending<-lm(Arbeitstabelle_top_10$SpendingProPerson~Arbeitstabelle_top_10$Zeit)
plot(Arbeitstabelle_top_10$Zeit, Arbeitstabelle_top_10$SpendingProPerson, ylab="Spending pro Person", xlab="Zeit")
```



Lineare Regression für unabhängige Variable "Zeit" und abhängige Variable "InternetAnteil" für top-10 Länder berechnen, Darstellung der Werte in Plot

Hide

```
lm.internet<-lm(Arbeitstabelle_top_10$InternetAnteil~Arbeitstabelle_top_10$Zeit)
plot(Arbeitstabelle_top_10$Zeit, Arbeitstabelle_top_10$InternetAnteil, ylab="Anteil von Internet", xlab="Zeit")
```



Zusammenfassung

Hide

```
summary(Arbeitstabelle_top_10)
```

```
Location   Zeit   SpendingProPerson InternetAnteil
AUT  :17  2006  :11   Min.   : 505.0   Min.   :19.05
BEL  :17  2007  :11   1st Qu.: 918.4   1st Qu.:64.73
CZE  :17  2008  :11   Median :1552.8   Median :80.00
DEU  :17  2009  :11   Mean    :3131.7   Mean    :74.28
DNK  :17  2010  :11   3rd Qu.: 5165.1   3rd Qu.:88.37
FIN  :17  2011  :11   Max.    :11539.2   Max.    :97.50
(Other):83 (Other):119
sqrt_InternetAnteil ZeitNum   LocationNum
Min.   :4.365   Min.   : 1.000   Min.   : 2.00
1st Qu.:8.046   1st Qu.: 5.000   1st Qu.:11.00
Median :8.945   Median : 9.000   Median :18.00
Mean   :8.537   Mean   : 9.005   Mean   :15.78
3rd Qu.:9.401   3rd Qu.:13.000   3rd Qu.:22.00
Max.   :9.874   Max.   :17.000   Max.   :28.00
```

Grafische Darstellungen top-10 Länder inklusive Deutschland

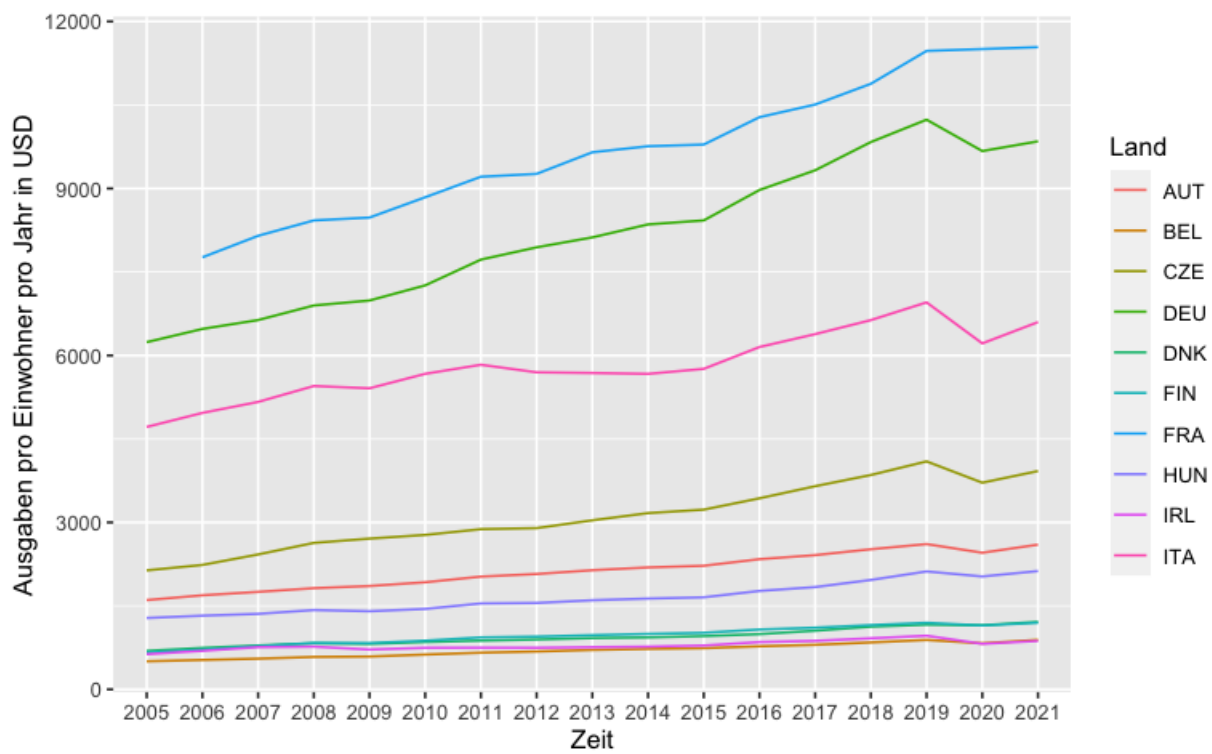
Hide

```
library(ggplot2)
library(RColorBrewer)

df <- Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<21 | Arbeitstabelle_top_10$LocationNum>21,c("SpendingProPerson", "ZeitNum", "Zeit", "LocationNum", "Location" )]

ggplot(df, aes(x = Zeit, y = SpendingProPerson, group = Location, col = Location))+

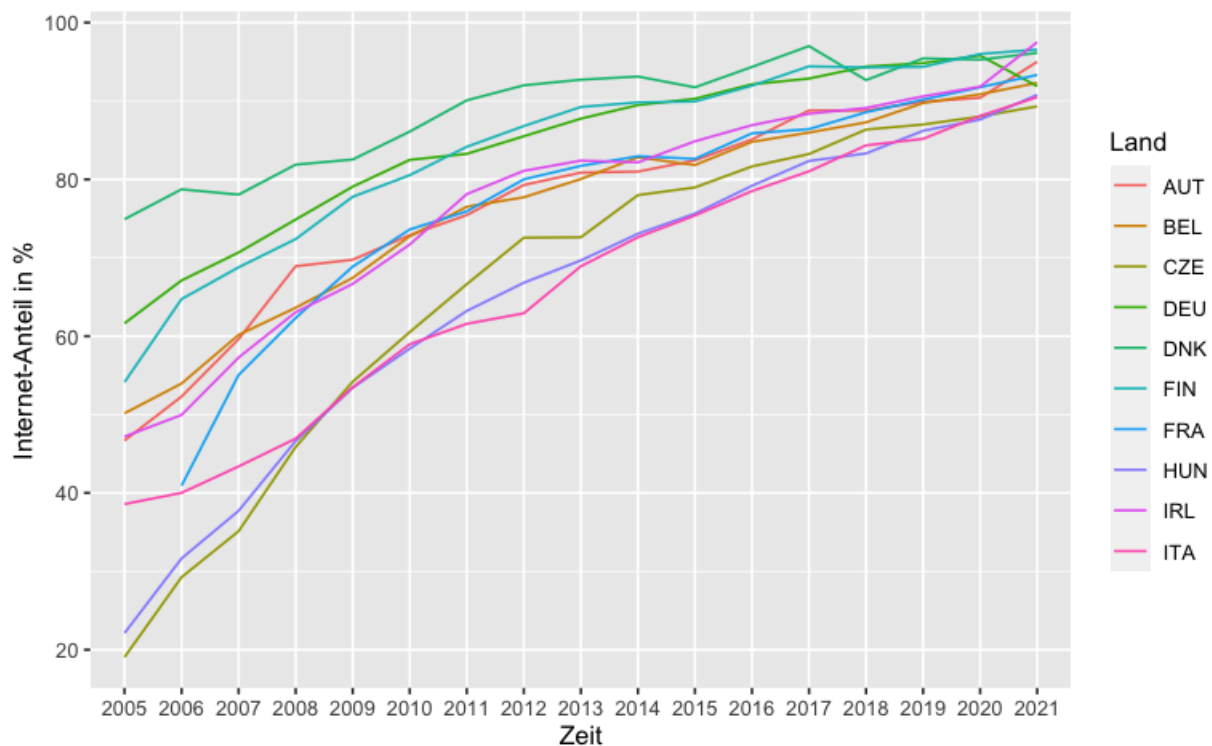
geom_line()+
labs(x="Zeit", y="Ausgaben pro Einwohner pro Jahr in USD", col="Land")
```



Hide

```
library(ggplot2)
library(RColorBrewer)
df <- Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<21 | Arbeitstabelle_top_10$LocationNum>21,c("InternetAnteil", "ZeitNum","Zeit", "
LocationNum", "Location" )]
ggplot(df, aes(x = Zeit, y = InternetAnteil, group = Location, col = Location))+

geom_line()+
labs(x="Zeit", y="Internet-Anteil in %", col="Land")
```



Hide

Berechnen der Parameter der Linearen Regression Erstellen neuer Dataframe

```
Arbeitstabelle_AUT_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=2, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_BEL_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=3 & Arbeitstabelle_top_10$LocationNum>2, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_CZE_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=11 & Arbeitstabelle_top_10$LocationNum>3, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_DEU_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=12 & Arbeitstabelle_top_10$LocationNum>11, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_DNK_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=13 & Arbeitstabelle_top_10$LocationNum>12, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_FIN_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=18 & Arbeitstabelle_top_10$LocationNum>13, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_FRA_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=19 & Arbeitstabelle_top_10$LocationNum>18, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_GRC_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=21 & Arbeitstabelle_top_10$LocationNum>19, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_HUN_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=22 & Arbeitstabelle_top_10$LocationNum>21, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_IRL_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=25 & Arbeitstabelle_top_10$LocationNum>22, c("ZeitNum", "SpendingProPerson")]
Arbeitstabelle_ITA_Sp<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=28 & Arbeitstabelle_top_10$LocationNum>25, c("ZeitNum", "SpendingProPerson")]
```

Hide

```
Arbeitstabelle_AUT_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=2, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_BEL_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=3 & Arbeitstabelle_top_10$LocationNum>2, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_CZE_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=11 & Arbeitstabelle_top_10$LocationNum>3, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_DEU_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=12 & Arbeitstabelle_top_10$LocationNum>11, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_DNK_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=13 & Arbeitstabelle_top_10$LocationNum>12, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_FIN_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=18 & Arbeitstabelle_top_10$LocationNum>13, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_FRA_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=19 & Arbeitstabelle_top_10$LocationNum>18, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_GRC_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=21 & Arbeitstabelle_top_10$LocationNum>19, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_HUN_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=22 & Arbeitstabelle_top_10$LocationNum>21, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_IRL_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=25 & Arbeitstabelle_top_10$LocationNum>22, c("ZeitNum", "sqrt_InternetAnteil")]
Arbeitstabelle_ITA_In<-Arbeitstabelle_top_10[Arbeitstabelle_top_10$LocationNum<=28 & Arbeitstabelle_top_10$LocationNum>25, c("ZeitNum", "sqrt_InternetAnteil")]
```