

2023-04-12

Anhang 1. Projekt DLBDBSC01

Kontrolle der Variablen

Hide

class(InternetAccess)

[1] "data.frame"

Hide

class(HouseholdSpending)

[1] "data.frame"

Hide

class(PopulationSize)

[1] "data.frame"

Hide

summary(InternetAccess)

LOCATION	INDICATOR	SUBJECT
Length:665	Length:665	Length:665
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

MEASURE	FREQUENCY	TIME
Length:665	Length:665	Min. :2005
Class :character	Class :character	1st Qu.:2009
Mode :character	Mode :character	Median :2013
		Mean :2013
		3rd Qu.:2018
		Max. :2022

Value	Flag.Codes
Min. : 7.659	Length:665
1st Qu.:60.500	Class :character
Median :77.990	Mode :character
Mean :72.155	
3rd Qu.:89.569	
Max. :99.932	

Hide

summary(HouseholdSpending)

LOCATION	INDICATOR	SUBJECT
Length:777	Length:777	Length:777
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

MEASURE	FREQUENCY	TIME
Length:777	Length:777	Min. :2005
Class :character	Class :character	1st Qu.:2009
Mode :character	Mode :character	Median :2013
		Mean :2013
		3rd Qu.:2017
		Max. :2021

Value	Flag.Codes
Min. : 6116	Mode:logical
1st Qu.: 105845	NA's:777
Median : 276242	
Mean : 1266282	
3rd Qu.: 1248403	
Max. :15902575	

[Hide](#)

summary(PopulationSize)

LOCATION	INDICATOR	SUBJECT
Length:35	Length:35	Length:35
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

MEASURE	FREQUENCY	TIME
Length:35	Length:35	Min. :2014
Class :character	Class :character	1st Qu.:2014
Mode :character	Mode :character	Median :2014
		Mean :2014
		3rd Qu.:2014
		Max. :2014

Value	Flag.Codes
Min. : 0.00	Mode:logical
1st Qu.:22.10	NA's:35
Median :41.50	
Mean :38.66	
3rd Qu.:54.70	
Max. :85.10	

Die relevante Spalten als Vektor/Datentyp speichern

[Hide](#)

```
Location<-as.factor(InternetAccess$LOCATION)
Zeit<-as.factor(InternetAccess$TIME)
InternetAnteil<-as.numeric(InternetAccess$Value)
Internetaccess01<-cbind.data.frame(Location, Zeit, InternetAnteil)
Internetaccess01
```

Location

<fctr>



BEL

BEL

BEL

BEL

BEL

BEL

BEL

BEL

BEL

BEL

1-10 of 665 rows | 1-1 of 3 columns

Previous **1** **2** Next

Hide

```
class(Internetaccess01)
```

```
[1] "data.frame"
```

Hide

```
summary(Internetaccess01)
```

```
Location   Zeit   InternetAnteil
AUT   : 18  2009  : 41  Min.   : 7.659
BEL   : 18  2012  : 41  1st Qu.:60.500
CZE   : 18  2008  : 39  Median :77.990
DEU   : 18  2010  : 39  Mean   :72.155
DNK   : 18  2013  : 39  3rd Qu.:89.569
ESP   : 18  2014  : 39  Max.   :99.932
(Other):557  (Other):427
```

Hide

```
Location<-as.factor(HouseholdSpending$LOCATION)
Zeit<-as.factor(HouseholdSpending$TIME)
SpendingAusgabenMinUsd<-as.numeric(HouseholdSpending$Value/4)#pro Person berechnen
HouseholdSpending01<-cbind.data.frame(Location, Zeit, SpendingAusgabenMinUsd)
HouseholdSpending01
```

Location

<fctr>



AUS

AUS

AUS

AUS

AUS

AUS

AUS

AUS

AUS

AUS

1-10 of 777 rows | 1-1 of 3 columns

Previous **1** **2** Next

Hide

class(HouseholdSpending01)

[1] "data.frame"

Hide

summary(HouseholdSpending01)

	Location	Zeit	SpendingAusgabenMinUsd
AUS	: 17	2005 : 46	Min. : 1529
AUT	: 17	2006 : 46	1st Qu.: 26461
BEL	: 17	2007 : 46	Median : 69061
CAN	: 17	2008 : 46	Mean : 316570
CHE	: 17	2009 : 46	3rd Qu.: 312101
CHL	: 17	2010 : 46	Max. : 3975644
(Other):	675	(Other):	501

Hide

```
Location<-as.factor(PopulationSize$LOCATION)
PopulationGroesse<-as.numeric(PopulationSize$Value)
PopulationSize01<-cbind.data.frame(Location, PopulationGroesse)
PopulationSize01
```

Location

<fctr>



POL

ESP

TUR

MEX

NOR

SVK

CHE

DNK

DEU

ISL

1-10 of 35 rows | 1-1 of 2 columns

Previous **1** **2** Next

[Hide](#)

```
class(PopulationSize01)
```

```
[1] "data.frame"
```

[Hide](#)

```
summary(PopulationSize01)
```

```
Location PopulationGroesse
AUS  : 1  Min.   : 0.00
AUT  : 1  1st Qu.:22.10
BEL  : 1  Median :41.50
CAN  : 1  Mean   :38.66
CHE  : 1  3rd Qu.:54.70
CHL  : 1  Max.   :85.10
(Other):29
```

Tabelle erstellen aus HouseholdSpending01 mit PopulationSize01

[Hide](#)

```
library(dplyr)
HouseholdSpendingNorm <- left_join(HouseholdSpending01, PopulationSize01, by = "Location")
HouseholdSpendingNorm
```

Location

<fctr>

AUS

AUS

AUS

AUS

AUS

AUS

AUS

AUS

AUS

AUS

1-10 of 777 rows | 1-1 of 4 columns

Previous **1** [2](#) [Next](#)

Nullwerte der Variable "Populationsgroesse" ausschließen

[Hide](#)

```
HouseholdSpendingNorm01 <- HouseholdSpendingNorm[HouseholdSpendingNorm$PopulationGroesse > 0, c("Location", "Zeit", "SpendingAusgaben", "MinUsd", "PopulationGroesse")]
HouseholdSpendingNorm01
```

1
2
3
4
5
6
7
8
9
10

1-10 of 726 rows | 1-1 of 4 columns

Previous **1** [2](#) [Next](#)

NA-Werte der Variable "Populationsgroesse" ausschließen

Hide

```
HouseholdSpendingNorm02 <- na.exclude(HouseholdSpendingNorm01)
HouseholdSpendingNorm02
```

1
2
3
4
5
6
7
8
9
10

1-10 of 544 rows | 1-1 of 4 columns

Previous **1** [2](#) [Next](#)

Ausgaben auf Populationsgroesse pro Land normalisieren

Hide

```
SpendingProPerson<-HouseholdSpendingNorm02$SpendingAusgabenMinUsd/HouseholdSpendingNorm02$PopulationGroesse
HouseholdSpendingNorm03<- cbind.data.frame(HouseholdSpendingNorm02, SpendingProPerson)
HouseholdSpendingNorm03
```

1
2
3
4
5
6
7
8
9
10

1-10 of 544 rows | 1-1 of 5 columns

Previous12Next

Zwei neue Arbeitstabellen "Arbeitstabelle01" und "Arbeitstabelle02" aus 2 Tabellen: "HouseholdSpendingNorm03" und InternetAccess01" erstellen

Hide

Arbeitstabelle01<-HouseholdSpendingNorm03

Hide

library(dplyr)
Arbeitstabelle <- left_join(HouseholdSpendingNorm03, Internetaccess01, by = c("Location", "Zeit"))
Arbeitstabelle

Location

<fctr>

AUS
AUS
AUS
AUS
AUS
AUS
AUS
AUS
AUS
AUS

1-10 of 544 rows | 1-1 of 6 columns

Previous12Next

NA-Werte ausschließen

Hide

Arbeitstabelle02<-na.exclude(Arbeitstabelle)
Arbeitstabelle02

1
2
3
4
6
8
10
12
18
19

1-10 of 482 rows | 1-1 of 6 columns

Previous12Next

Ergebnis: Arbeitstabelle01 und Arbeitstabelle02 erstellt