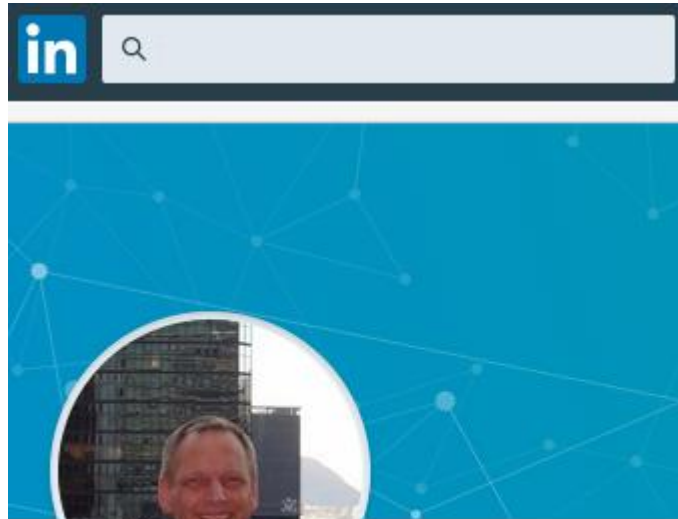




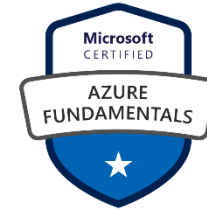
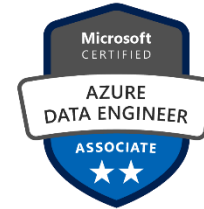
PowerBi & Korrelation

about me

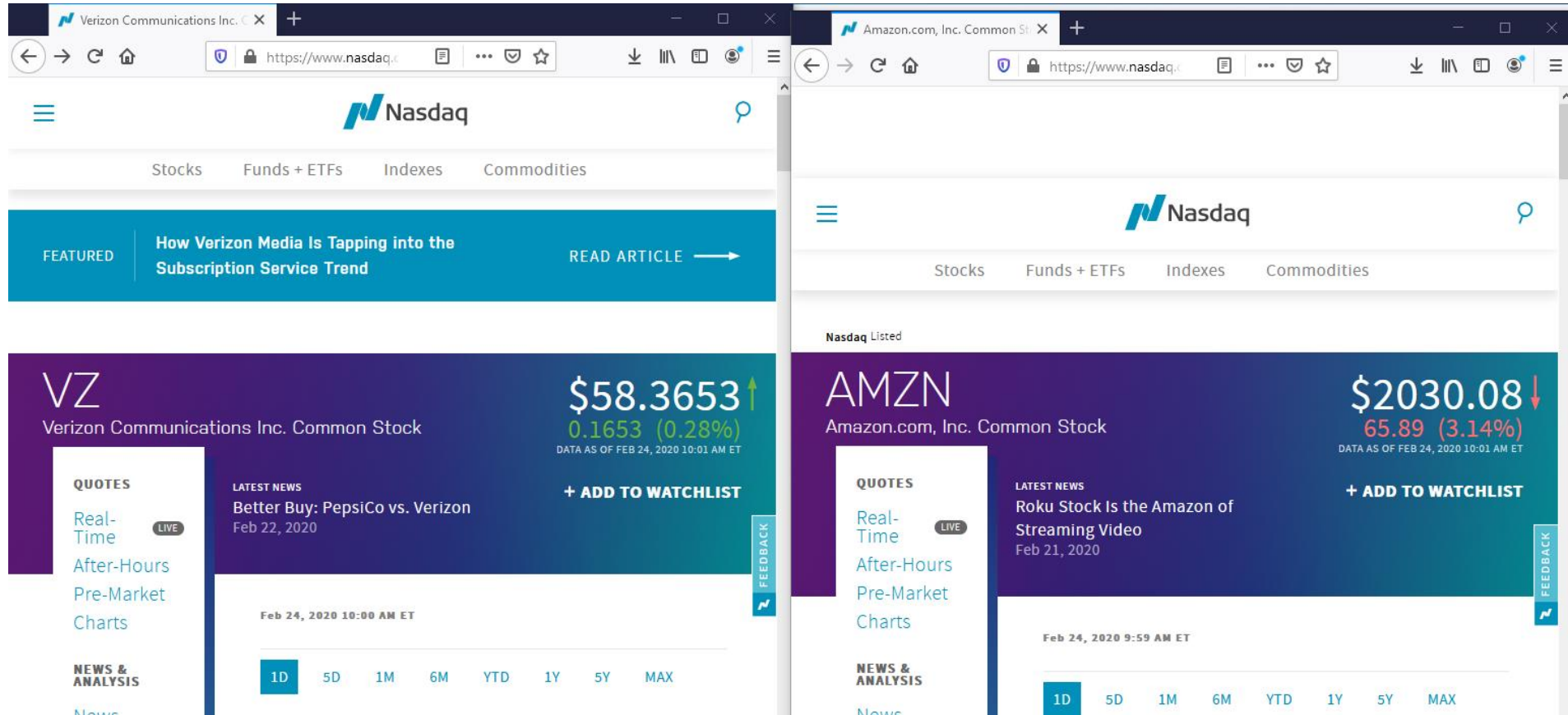


Alexander Karl

<https://www.linkedin.com/in/alexander-karl-44561012a/>



Fragestellungen ...



<https://www.nasdaq.com/market-activity/stocks/amzn>



Fragestellungen ...

ist das Geschäft/ der Geschäftsbereich
"besser" als ...

in bezug auf systematisches und unsystematisches Risiko



Fragestellungen ...

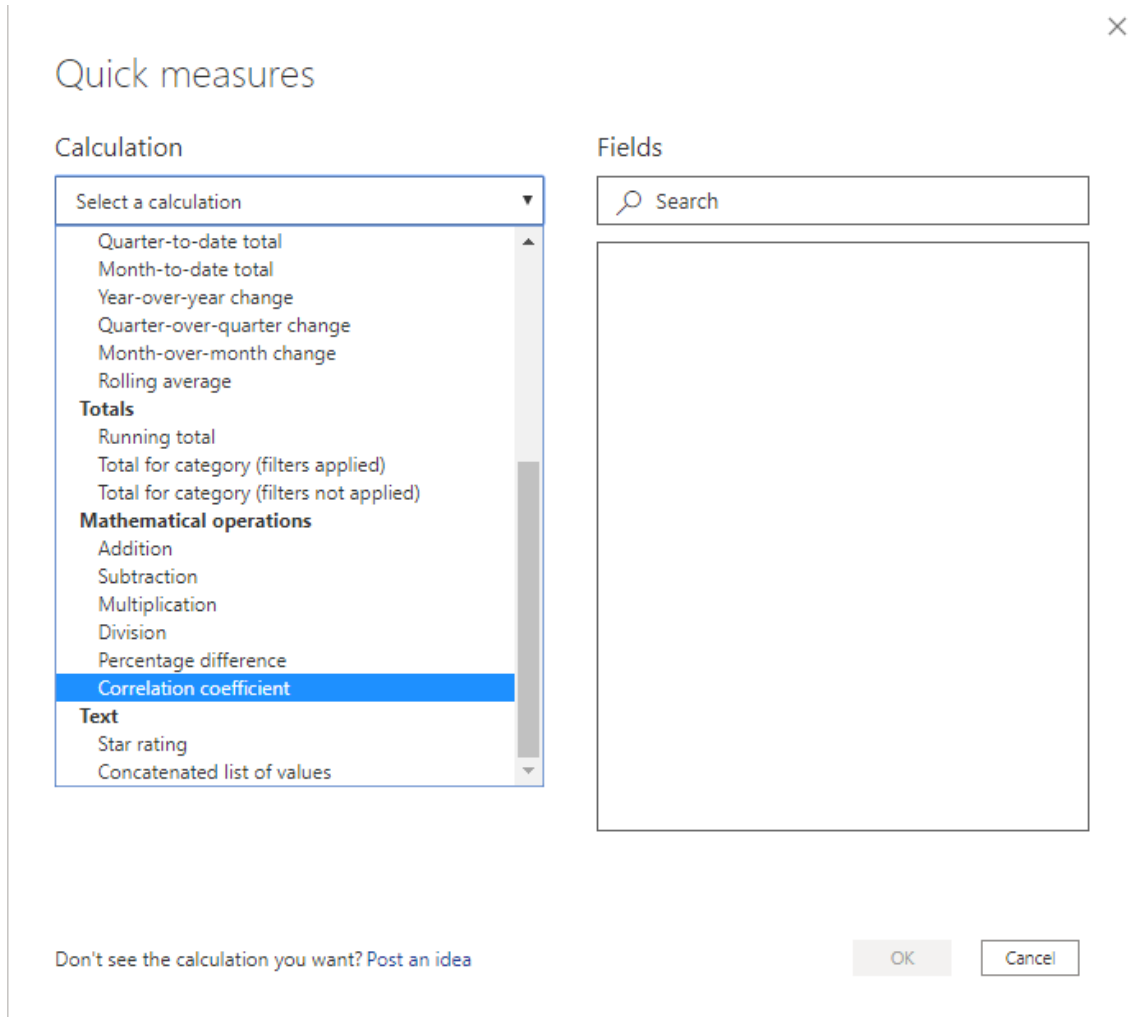
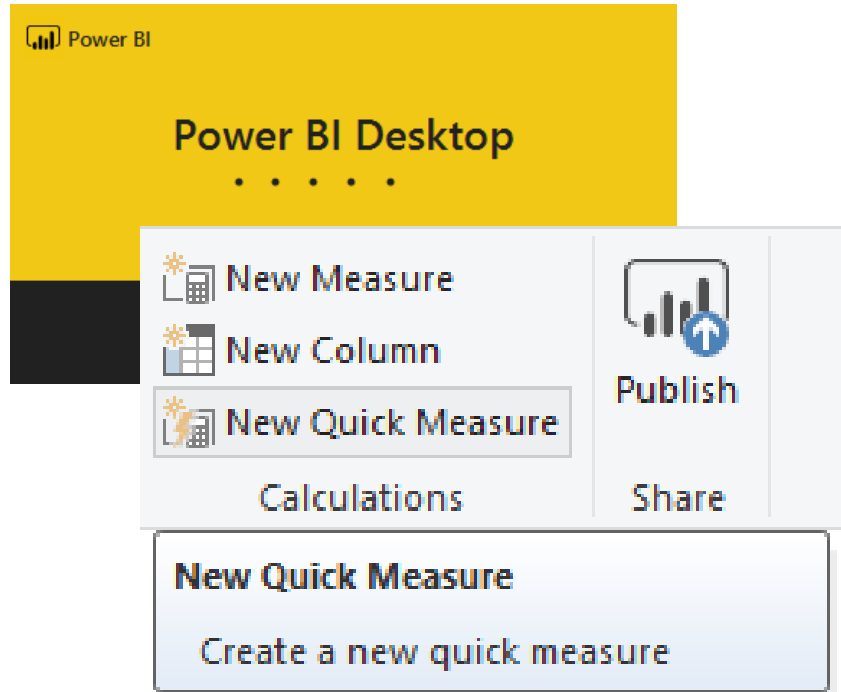
ist das Geschäft/ der Geschäftsbereich
"besser" als ...

bisher mit Punkt-Vergleich (meist Zeit-Punkt)

Jetzt mit Reihen-Vergleich (meist Zeit-Reihen)



PowerBI Quick Measures



<https://docs.microsoft.com/de-de/power-bi/transform-model/desktop-quick-measures>

<https://community.powerbi.com/t5/Quick-Measures-Gallery/Correlation-coefficient/m-p/196274>





Demo

Clipboard: Paste, Cut, Copy, Format Painter

External data: Get Data, Recent Sources, Enter Data, Edit Queries, Refresh

Insert: New Page, New Visual, Ask A Question, Buttons, Text box, Image, Shapes

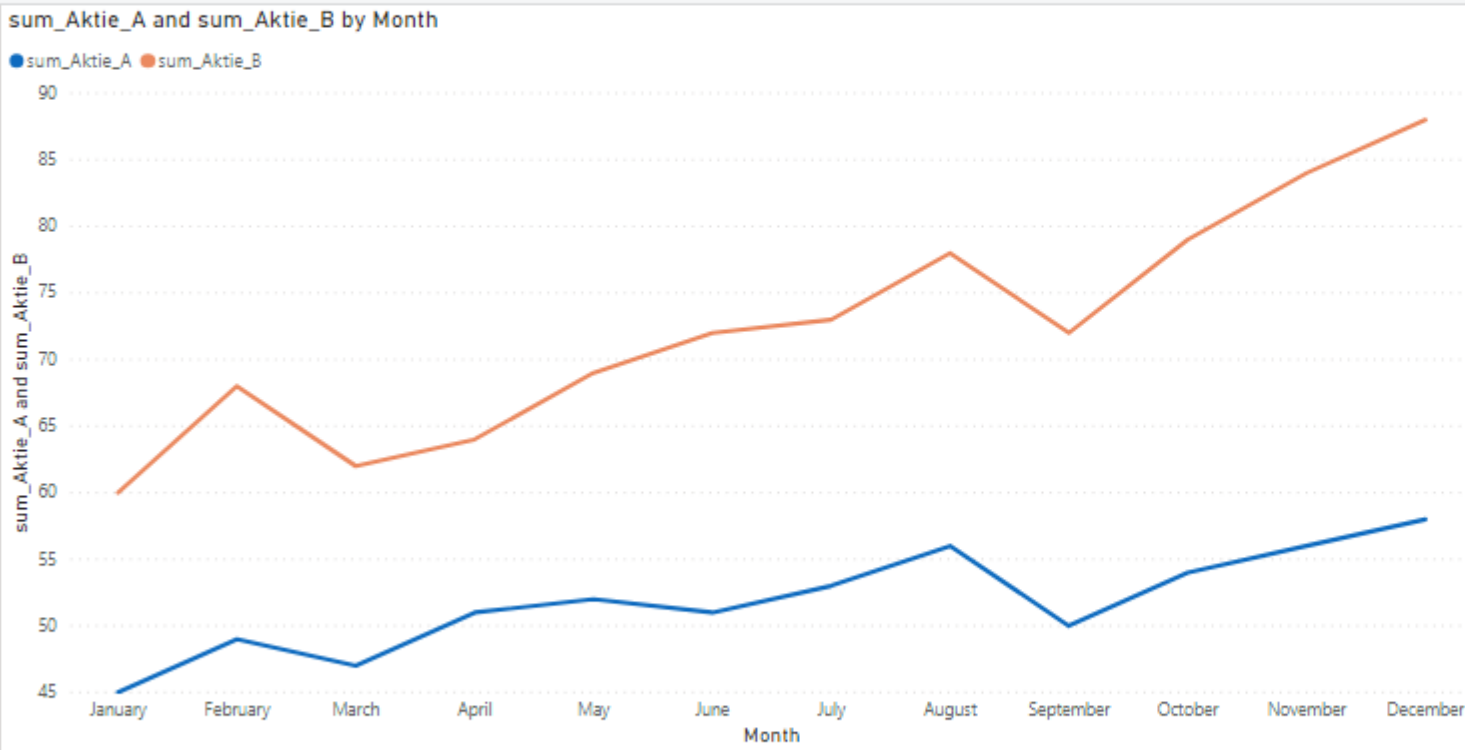
Custom visuals: From Marketplace, From File

Themes: Switch Theme

Relationships: Manage Relationships

Calculations: New Measure, New Column, New Quick Measure

Share: Publish



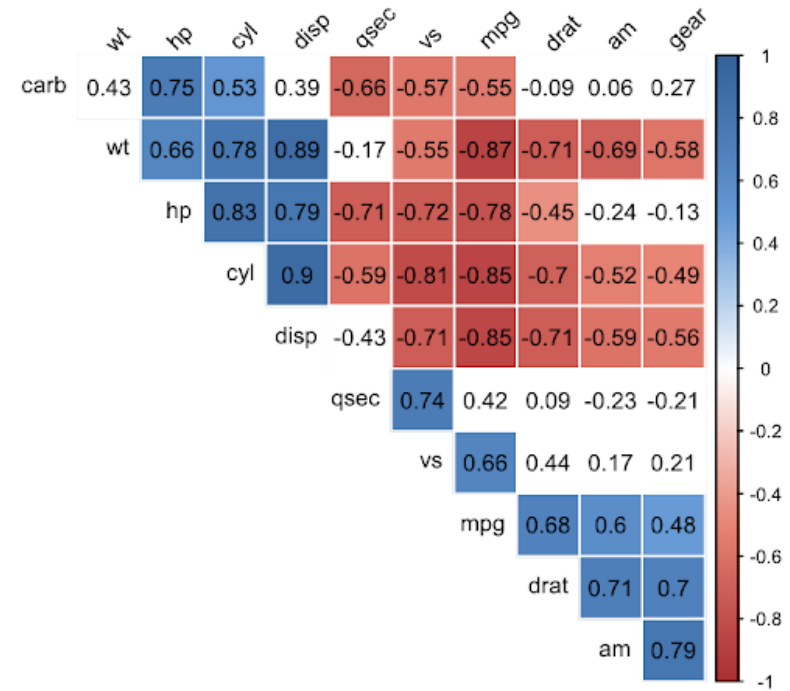
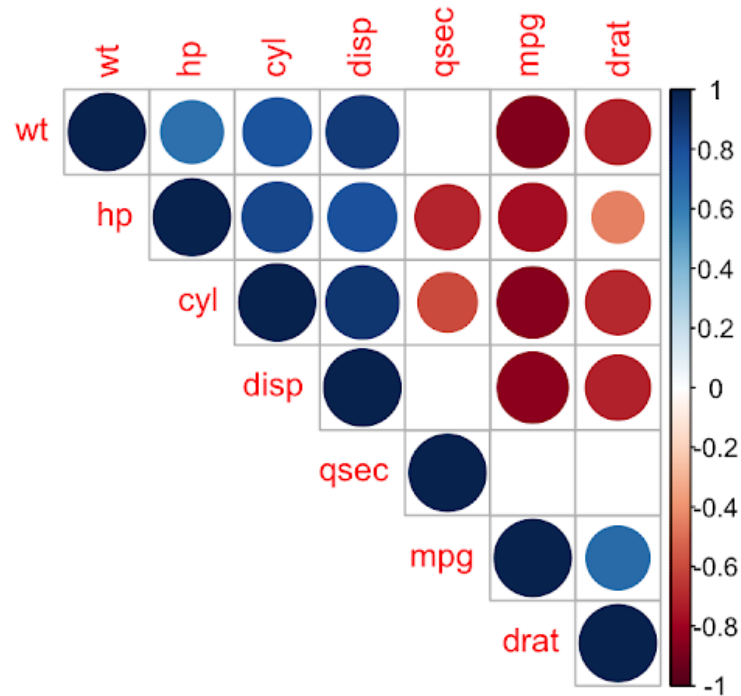
0,93
corr_A_B

Fields

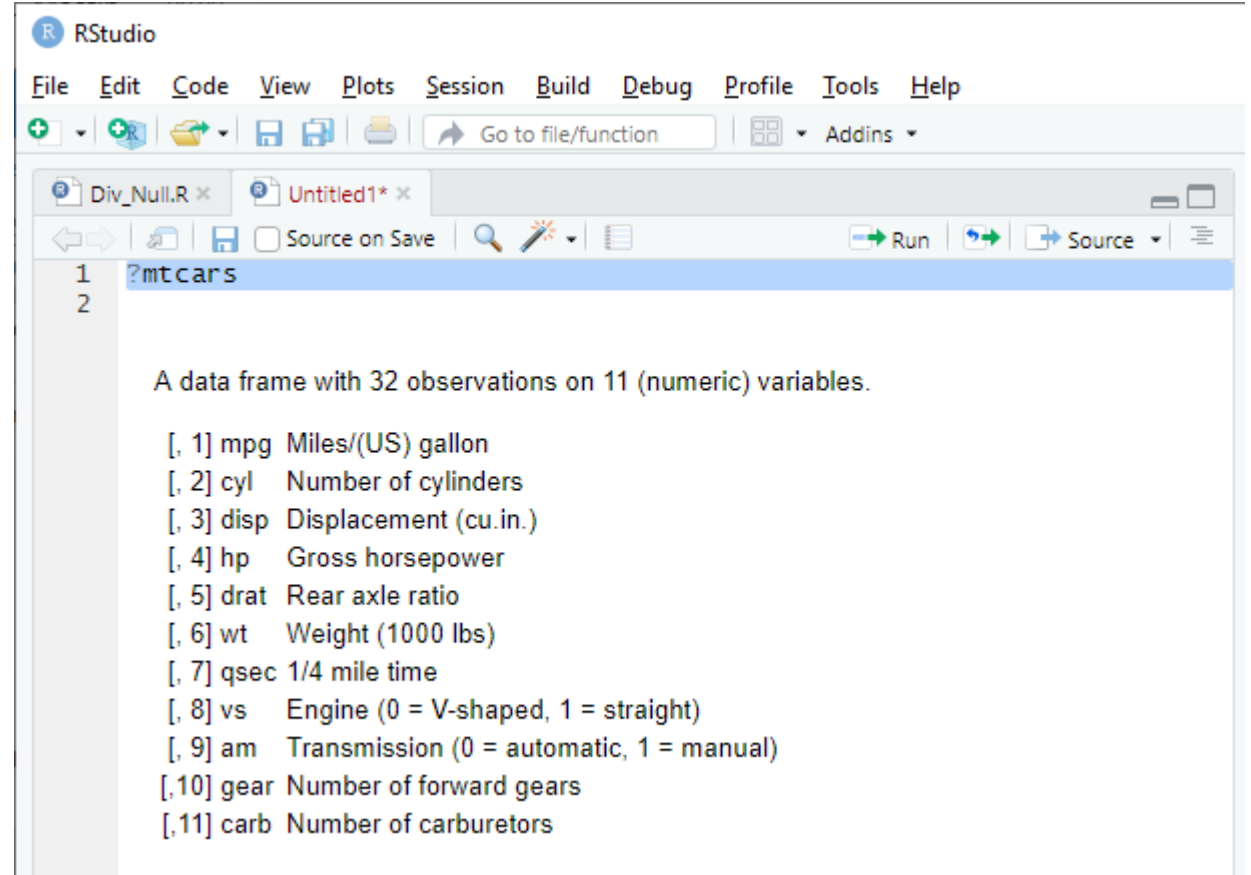
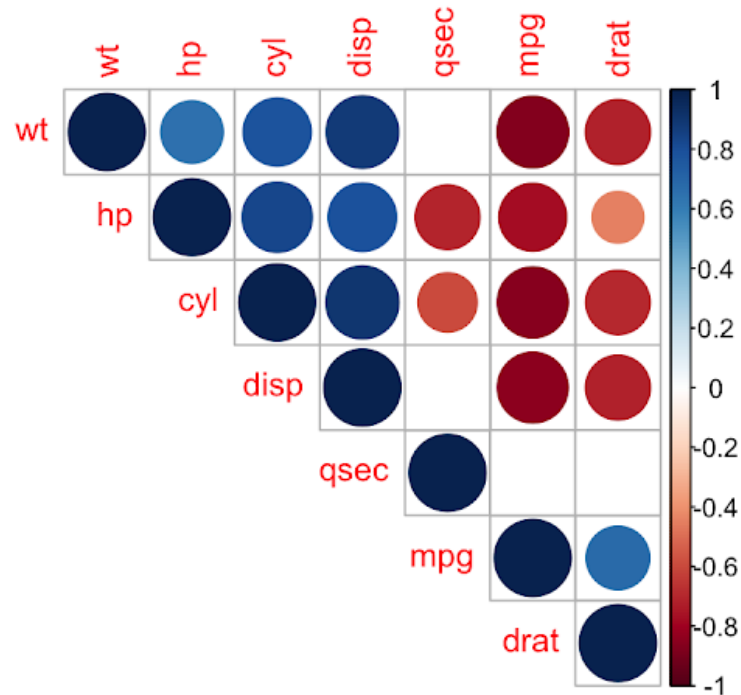
Search

- Visualizations
- step01
 - Aktie_A
 - Aktie_B
 - corr_A_B
 - Monat
 - sum_Aktie_A
 - sum_Aktie_B

Ausblick mit



Ausblick mit



RStudio interface showing the help page for the 'mtcars' dataset. The code editor contains the command `?mtcars`. The output displays the following information:

```
1 ?mtcars
2
A data frame with 32 observations on 11 (numeric) variables.

 [, 1] mpg Miles/(US) gallon
 [, 2] cyl Number of cylinders
 [, 3] disp Displacement (cu.in.)
 [, 4] hp Gross horsepower
 [, 5] drat Rear axle ratio
 [, 6] wt Weight (1000 lbs)
 [, 7] qsec 1/4 mile time
 [, 8] vs Engine (0 = V-shaped, 1 = straight)
 [, 9] am Transmission (0 = automatic, 1 = manual)
[ ,10] gear Number of forward gears
[ ,11] carb Number of carburetors
```



Scalar- and Aggregate-Functions

The image shows two overlapping screenshots of the Microsoft Excel interface, illustrating the use of scalar and aggregate functions.

Left Screenshot (Worksheet 1):

- Formula Bar:** Shows the formula `=MONTH(C2)`.
- Worksheet Data:**

ID	Umsatz	Verkaufsdatum	VerkaufsMonat
1	200	18.08.2020	<code>=MONTH(C2)</code>
2	150	19.08.2020	<code>MONTH(serial_num)</code>
3	220	20.08.2020	8
4	300	21.08.2020	8
5	240	24.08.2020	8
6	200	25.08.2020	8
7	180	26.08.2020	8
8	220	27.08.2020	8
9	300	28.08.2020	8
10	240	31.08.2020	8
11	200	01.09.2020	9
12	180	02.09.2020	9
13	240	03.09.2020	9
14	200	04.09.2020	9
15	180	07.09.2020	9

Right Screenshot (Worksheet 2):

- Formula Bar:** Shows the formula `=SUM(B2:B16)`.
- Worksheet Data:**

ID	Umsatz	Verkaufsdatum	VerkaufsMonat
1	200	18.08.2020	8
2	150	19.08.2020	8
3	220	20.08.2020	8
4	300	21.08.2020	8
5	240	24.08.2020	8
6	200	25.08.2020	8
7	180	26.08.2020	8
8	220	27.08.2020	8
9	300	28.08.2020	8
10	240	31.08.2020	8
11	200	01.09.2020	9
12	180	02.09.2020	9
13	240	03.09.2020	9
14	200	04.09.2020	9
15	180	07.09.2020	9

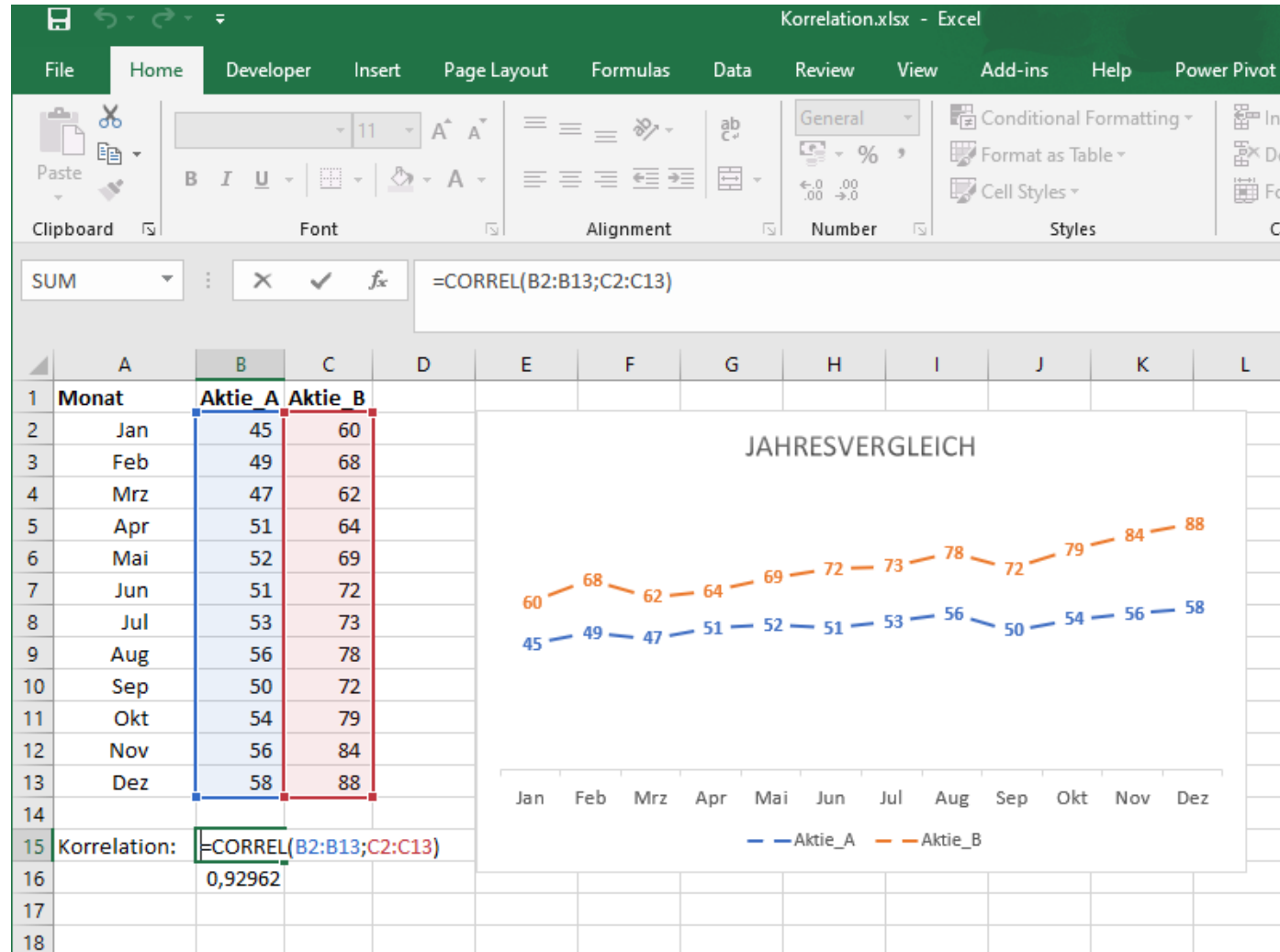
Bottom Screenshot (Worksheet 3):

- Formula Bar:** Shows the formula `=SUM(B2:B16)`.
- Worksheet Data:**

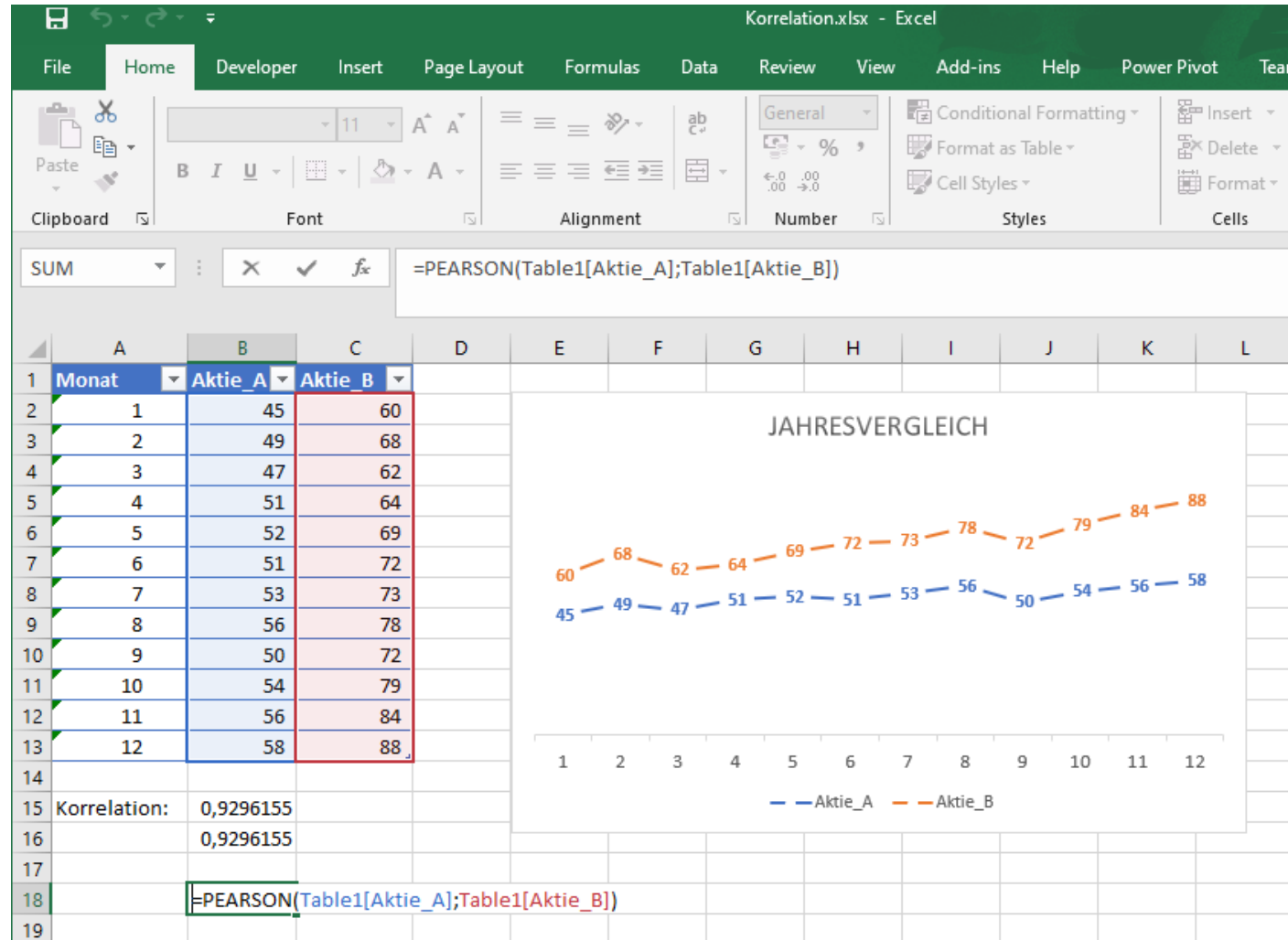
ID	Umsatz	Verkaufsdatum	VerkaufsMonat
1	200	18.08.2020	8
2	150	19.08.2020	8
3	220	20.08.2020	8
4	300	21.08.2020	8
5	240	24.08.2020	8
6	200	25.08.2020	8
7	180	26.08.2020	8
8	220	27.08.2020	8
9	300	28.08.2020	8
10	240	31.08.2020	8
11	200	01.09.2020	9
12	180	02.09.2020	9
13	240	03.09.2020	9
14	200	04.09.2020	9
15	180	07.09.2020	9



Statistical-Functions



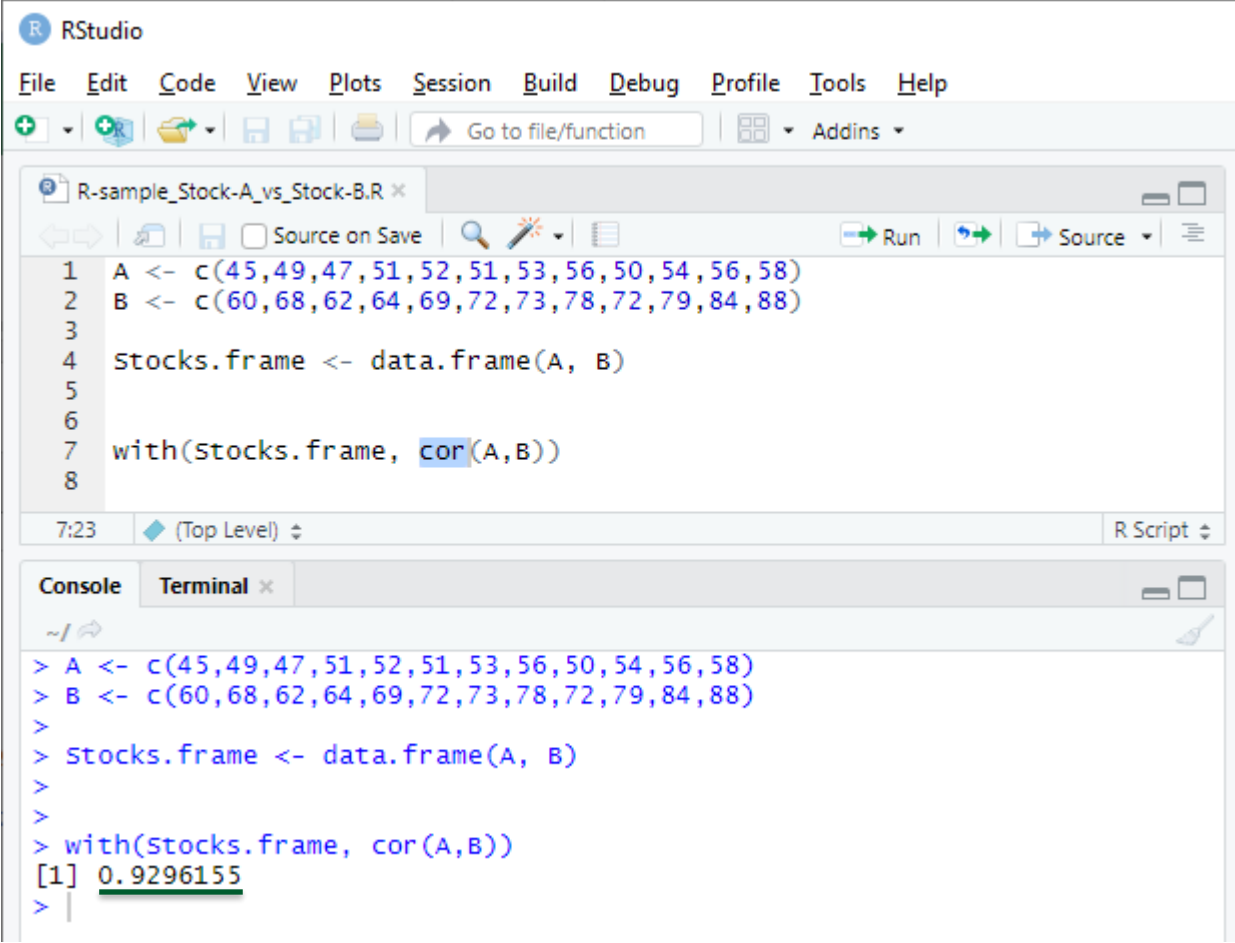
Statistical-Functions



<https://support.microsoft.com/en-us/office/pearson-function-0c3e30fc-e5af-49c4-808a-3ef66e034c18>



Solution with and R-Studio



The screenshot shows the RStudio environment. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with icons for file operations and a search bar. The main editor window displays the following R code:

```
1 A <- c(45,49,47,51,52,51,53,56,50,54,56,58)
2 B <- c(60,68,62,64,69,72,73,78,72,79,84,88)
3
4 Stocks.frame <- data.frame(A, B)
5
6
7 with(Stocks.frame, cor(A,B))
8
```

The console window at the bottom shows the execution of the code, resulting in the correlation coefficient 0.9296155:

```
> A <- c(45,49,47,51,52,51,53,56,50,54,56,58)
> B <- c(60,68,62,64,69,72,73,78,72,79,84,88)
>
> Stocks.frame <- data.frame(A, B)
>
>
> with(Stocks.frame, cor(A,B))
[1] 0.9296155
> |
```



Grundlagen Statistik

von den Daten zur Formel ...

Daten

Skalenniveau

Lage

Parameter

Streuung

Maße (var, sd)



Skalierungen

- Nominalskala → Cramer's V
 - Geschlecht, Haarfarbe
- Ordinalskala → Spearman
 - Bildung einer Rangfolge ist möglich
- Kardinalskala → Pearson
 - Gewicht, Einkommen, etc.

<https://studyflix.de/statistik/skalenniveaus-1050>



Lageparameter

Überblick der Lageparameter

Arithmetisches Mittel

(auch: Mittelwert oder Durchschnittswert)

(Average)

→ statistischer Durchschnittswert

→ Zur Berechnung addieren wir alle Beobachtungsdaten und teilen dann die Summe durch die Anzahl der Daten.

Beispiel

Die Körpergröße von acht Personen (in cm): 150, 155, 160, 165, 170, 175, 180, 185

Arithmetisches Mittel: 167.5

Median

→ Wert, der genau in der Mitte einer Datenreihe liegt, die nach der Größe geordnet ist

Beispiel

geordnete Datenreihe: 10, 20, 30, 40, 50, 60, 70

Median: 40



Streuungsmaße

Varianz & Standardabweichung

Varianz_Standard-Abweichung.xlsx - Excel

File Home Developer Insert Page Layout Formulas Data Review View Add-ins Help Power Pivot

Clipboard Font Alignment Number Conditional Formatting

	A	B	C	D	E	F	G	H
1		Abweichung			Abweichung	Quad. Abweichung		
2	Größe Gruppe A			Größe Gruppe B				
3	122 cm	0 cm		127 cm	5 cm	25 cm ²		
4	122 cm	0 cm		120 cm	-2 cm	4 cm ²		
5	122 cm	0 cm		132 cm	10 cm	100 cm ²		
6	122 cm	0 cm		117 cm	-5 cm	25 cm ²		
7	122 cm	0 cm		114 cm	-8 cm	64 cm ²		
8								
9	122 cm	0 cm	SUM	122 cm	0 cm	218 cm ²	Varianz	
10		0 cm	AVERAGE		0 cm	14,8 cm	Standardabweichung	
11								
12	Null-Eigenschaft des arithmetischen Mittels.							
13								
14								



Babe Ruth vs. Barry Bonds



Wer ist erfolgreicher ?

Babe Ruth erzielte 1927 60 Homeruns,
Barry Bonds im Jahr 2001 73 Homeruns ?

präziser ...

Wer war "mit Abstand" besser ?



Babe Ruth vs. Barry Bonds



Jahr	1927	2001
Besten 50		
Durchschnitt	12,68 Homeruns	37,02 Homeruns
Standardabweich.	10,49	9,64
	$\frac{60 - 12,68}{10,49}$	$\frac{73 - 37,02}{9,64}$
Z-Wert	4,51	3,73



Formel Korrelation

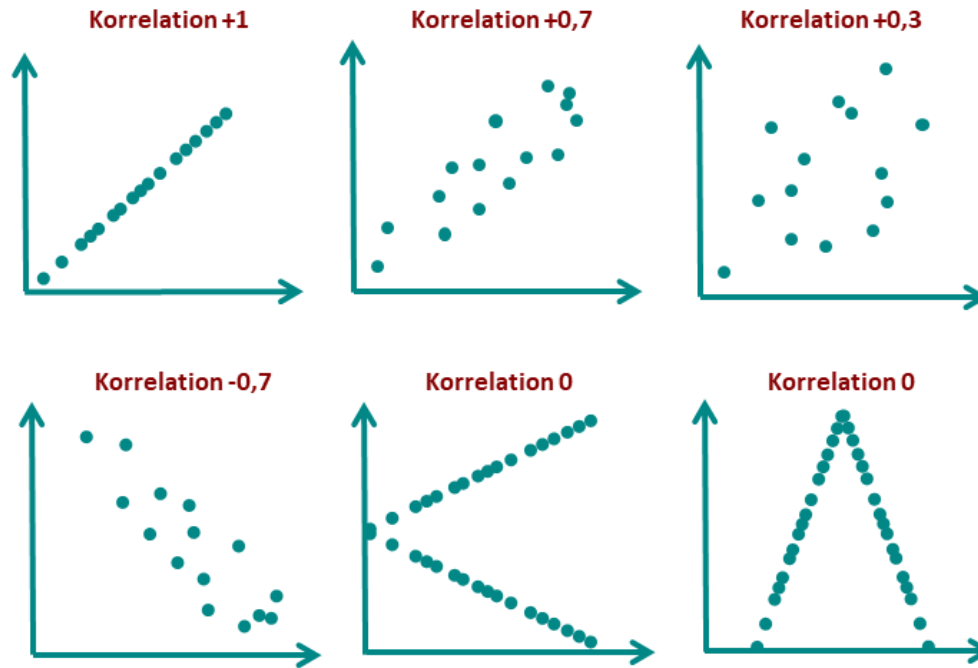
Formel zum Korrelationskoeffizient nach Pearson

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(N - 1) s_x s_y} = \frac{s_{xy}}{s_x s_y}$$

r	Korrelationskoeffizient
x_i	Beobachtungswerte der Variable x
y_i	Beobachtungswerte der Variable y
\bar{x}	Arithmetisches Mittel aller Wert von x
\bar{y}	Arithmetisches Mittel aller Wert von y
N	Gesamtanzahl
s_{xy}	Kovarianz der Variablen x und y
s_x	Standardabweichung der Variable x
s_y	Standardabweichung der Variable y



Interpretation der Ergebnisse



Betrag von r	Stärke des Zusammenhangs
$0,0 < 0,1$	kein Zusammenhang
$0,1 < 0,3$	geringer Zusammenhang
$0,3 < 0,5$	mittlerer Zusammenhang
$0,5 < 0,7$	hoher Zusammenhang
$0,7 < 1$	sehr hoher Zusammenhang

<https://datatab.de/tutorial/korrelation>



weitere statistische Anmerkungen

griechische vs.

lateinische Buchstaben

>> **Grundgesamtheit**

Stichprobe

Varianz:

>>
$$\sigma^2 = \frac{\sum(X - \bar{X})^2}{N}$$

$$s^2 = \frac{\sum(X - \bar{X})^2}{N-1}$$

Standardabweichung:

>>
$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(X - \bar{X})^2}{N}}$$

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(X - \bar{X})^2}{N-1}}$$



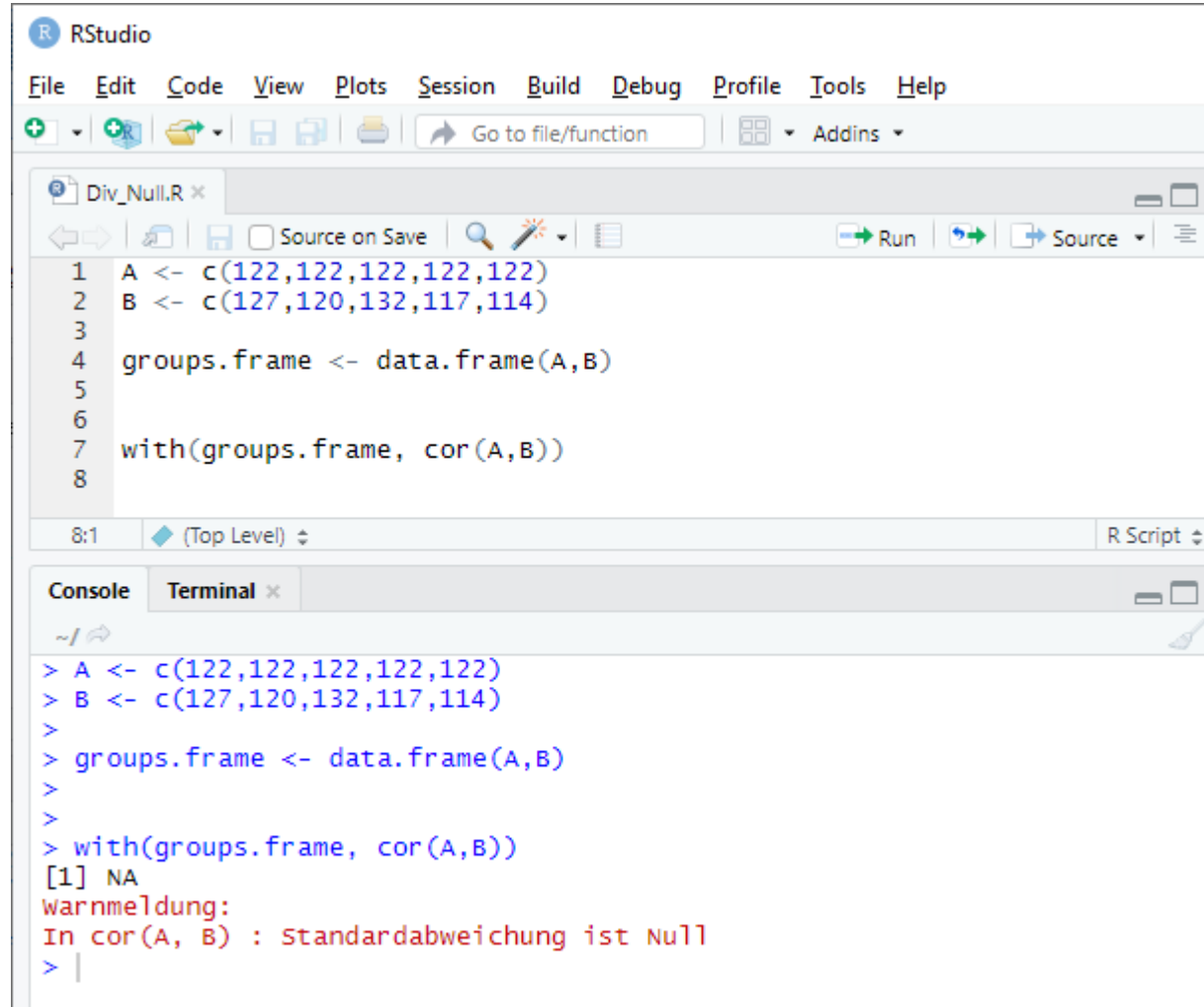
Besonderheiten >> #Div/0!

The screenshot shows the Microsoft Excel interface with the following data and error:

	A	B	C	D	E	F	G
1		Abweichung			Abweichung	Quad. Abweichung	
2	Größe Gruppe A			Größe Gruppe B			
3	122 cm	0 cm		127 cm	5 cm	25 cm ²	
4	122 cm	0 cm		120 cm	-2 cm	4 cm ²	
5	122 cm	0 cm		132 cm	10 cm	100 cm ²	
6	122 cm	0 cm		117 cm	-5 cm	25 cm ²	
7	122 cm	0 cm		114 cm	-8 cm	64 cm ²	
15	=CORREL(A3:A7;D3:D7)						
16	CORREL(array1; array2)			#DIV/0!			



Besonderheiten >> #Div/0!



The screenshot displays the RStudio interface. The script editor contains the following R code:

```
1 A <- c(122,122,122,122,122)
2 B <- c(127,120,132,117,114)
3
4 groups.frame <- data.frame(A,B)
5
6
7 with(groups.frame, cor(A,B))
8
```

The console output shows the execution of the code and the resulting error:

```
> A <- c(122,122,122,122,122)
> B <- c(127,120,132,117,114)
>
> groups.frame <- data.frame(A,B)
>
>
> with(groups.frame, cor(A,B))
[1] NA
warnmeldung:
In cor(A, B) : standardabweichung ist Null
> |
```



Besonderheiten

The screenshot displays the Power BI Desktop interface with the following components:

- Top Ribbon:** Includes tabs for Date!, Home, View, Modeling, Help, Format, and Data / Drill. The 'Visual tools' section is active, showing options like New Measure, New Column, and New Quick Measure.
- Main Canvas:** A line chart with two data series (blue and red) plotted against a date axis from 2017-11-06 to 2019-07-31. The y-axis ranges from 2400 to 3000. A green horizontal line is drawn at approximately y=2850.
- Right Panel:** Contains the 'Filters' pane (with 'Date - Month is (All)') and the 'Visualizations' pane (with 'Date' and 'Month' selected in the Values list).
- Fields Pane:** Shows a hierarchy for 'Date' with sub-items for Year, Quarter, Month, and Day.
- Error Dialog:** A modal window titled 'Can't display this visual.' with the message: 'No image was created. The R code didn't result in creation of any visuals. Make sure your R script results in a plot to the R default device.' It includes a 'See details' link and a 'Copy details' button.
- Bottom Bar:** Labeled 'R script editor'.



Berechnungen für Measure-“Kombinationen“



W Gaußsche Summenformel – Wi X +

https://de.wikipedia.org/wiki/Gaußsche_Summenformel



Man braucht nun nur mehr die Anzahl $n \cdot (n + 1)$ aller Kästchen zu halbieren, was sofort zur gesuchten Anzahl $\frac{n \cdot (n + 1)}{2}$ der grünen Kästchen führt.

Herkunft der Bezeichnung [\[Bearbeiten | Quelltext bearbeiten \]](#)

Diese Summenformel wie auch die Summenformel für die ersten n Quadratzahlen war bereits in der vorgriechischen Mathematik bekannt.

[Carl Friedrich Gauß](#) entdeckte diese Formel als neunjähriger Schüler wieder. Die Geschichte ist durch [Wolfgang Sartorius von Waltershausen](#) überliefert:

„Der junge Gauss war kaum in die Rechenklasse eingetreten, als Büttner die Summation einer arithmetischen Reihe aufgab. Die Aufgabe war indess kaum ausgesprochen als Gauss die Tafel mit den im niedern Braunschweiger Dialekt gesprochenen Worten auf den Tisch wirft: »Ligget se'.« (Da liegt sie.)“

– [WOLFGANG SARTORIUS VON WALTERSHAUSEN](#)^[1]



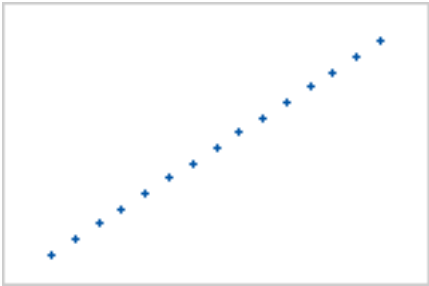
Pearson vs. Spearman's Rank Correlation abhängig vom Skalenniveau

- Nominalskala → Cramer's V
 - Geschlecht, Haarfarbe
- Ordinalskala → Spearman
 - Bildung einer Rangfolge ist möglich
- Kardinalskala → Pearson
 - Gewicht, Einkommen, etc.

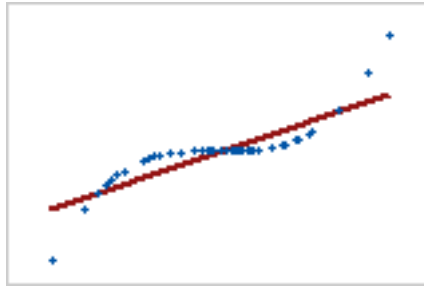
<https://studyflix.de/statistik/skalenniveaus-1050>



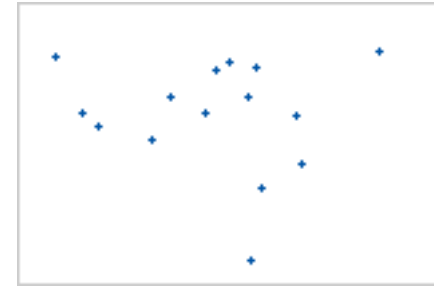
Pearson vs. Spearman's Rank Correlation



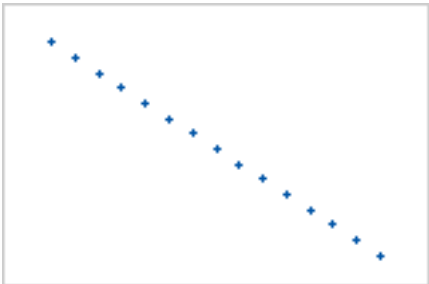
Pearson = +1, Spearman = +1



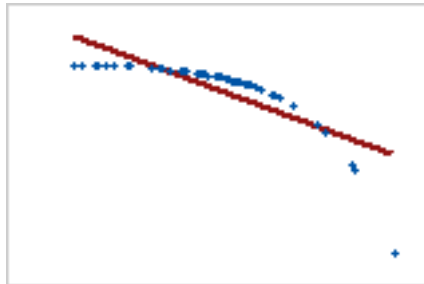
Pearson = +0,851, Spearman = +1



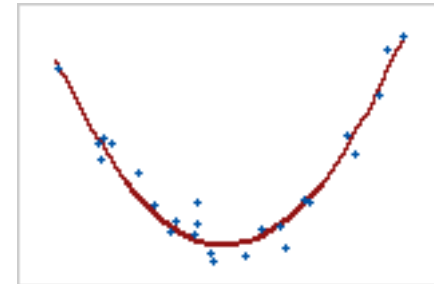
Pearson = -0,093, Spearman = -0,093



Pearson = -1, Spearman = -1



Pearson = -0,799, Spearman = -1



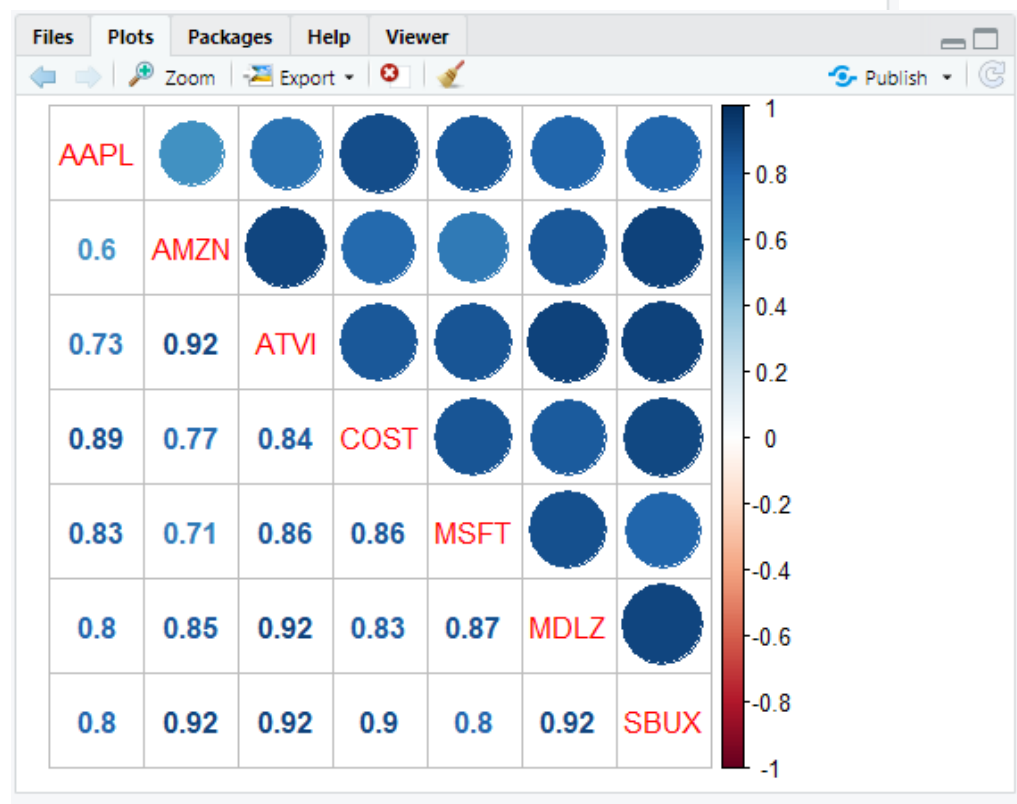
Koeffizient von 0





Demo

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ + + + + Go to file/function Addins
Source on Save Run Source
1 ##install.packages('RODBC')
2 library(RODBC)
3 ##install.packages('corrplot')
4 library(corrplot)
5
6 dbhandle <- odbcDriverConnect("driver={SQL Server};server=localhost;database=eodData;trusted_connection=true")
7 sqltab <- sqlQuery(dbhandle, "select [AAPL],[AMZN],[ATVI],[COST],[MSFT],[MDLZ],[SBUX] from dbo.corrTable")
8
9 M <- cor(sqltab) ## M like Matrix
10
11 ##-----
12 ## is a wrapped function for mixed visualization style.
13 corrplot.mixed(N)
14
15
16 ##-----
```



Zusammenfassung:

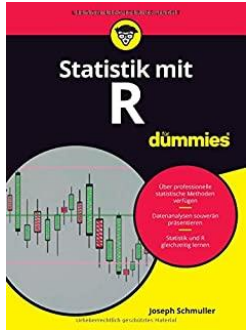
PowerBI Quick Measures (limitiert auf 2 Measures)

Excel (limitiert auf 2 arrays)

R integration (sehr gute graphische Umsetzung)



weitere links & Empfehlungen



https://www.amazon.de/Statistik-mit-R-f%C3%BCr-Dummies/dp/3527713980/ref=sr_1_5

<https://www.scribbr.de/statistik/korrelation/>

<https://www.scribbr.de/statistik/korrelationskoeffizient/>

<https://www.scribbr.de/statistik/lageparameter/>

<https://www.scribbr.de/statistik/streuungsma%C3%9F/>

<https://datatab.de/tutorial/korrelation>

<http://www.sthda.com/english/wiki/visualize-correlation-matrix-using-correlogram>

<https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html>

<https://de.excel-translator.de/mittelwert/>

<https://support.microsoft.com/en-us/office/excel-functions-alphabetical-b3944572-255d-4efb-bb96-c6d90033e188>



Korrelationsanalyse - Datatab

https://datatab.de/tutorial/korrelation

Statistik Rechner Diagramme Tutorial Statistik-Kurs Premium FAQ Login EN DE

→ Hier geht's zum **Online Statistik Rechner** ←


Theorie Beispiel

Korrelationsanalyse

Mithilfe von **Korrelationsanalysen** kann der lineare Zusammenhang von Variablen untersucht werden, weshalb diese auch als **Zusammenhangsanalysen** bezeichnet werden. Wie stark die **Korrelation** ist, ergibt sich über den **Korrelationskoeffizienten**, der von -1 bis +1 schwankt. Damit kann mit der Korrelationsanalysen eine Aussage über die Stärke und die Richtung des Zusammenhangs gemacht werden.

Korrelationsanalyse online berechnen ... Später ans... Teilen

Korrelationsanalyse online berechnen mit DATAtab



DATAtab
Die Webapp für Datenanalyse.



.. bis zum nächsten

