

Schatzsuche –

auf der Suche nach neuen Zusammenhängen:
Lessons learned aus dem Unternehmensalltag

Dr. Mihael Ankerst
Allianz Deutschland AG
19.2.2019



Inhalt des Vortrags

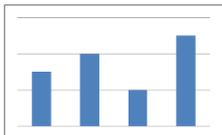
Sie werden u.a. folgendes mitnehmen...



Was ist der kürzeste Weg einer Ameise?

$$\bar{x} < \min x_i$$

Wann ist der Mittelwert einer Größe kleiner als alle Ausprägungen dieser Größe?



Was kann man mit einem Balkendiagramm so alles machen?



Allianz - 1

Geschäftsergebnis Allianz Deutschland 2017

Kunden

20,4 Mio.

zum Jahresende
2017

Marktanteil

17,4%

am deutschen
Versicherungs-
markt

Leistungen
an Kunden

18,9 Mrd.

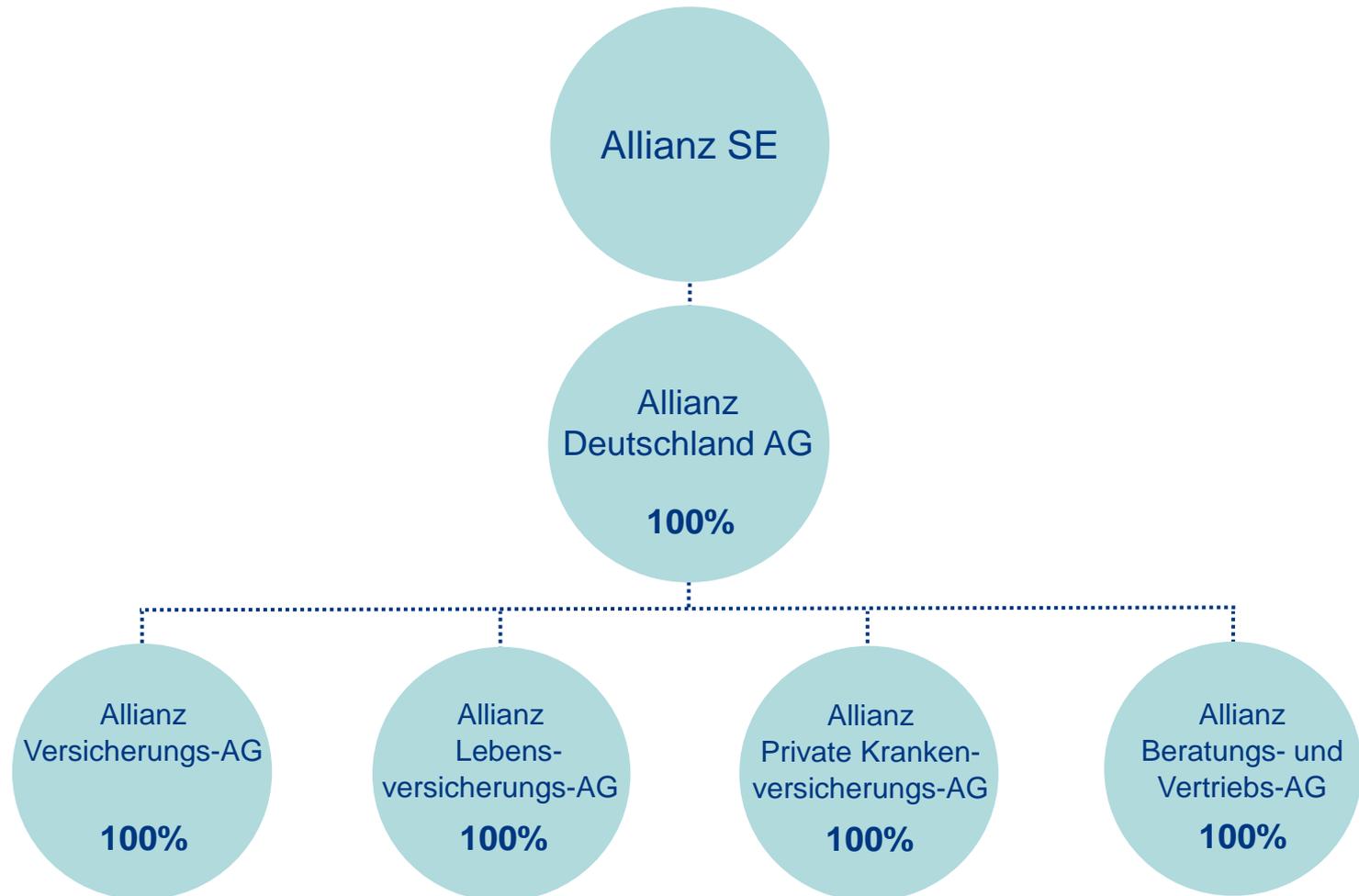
EUR





Allianz - 2

Organisationsstruktur der Allianz





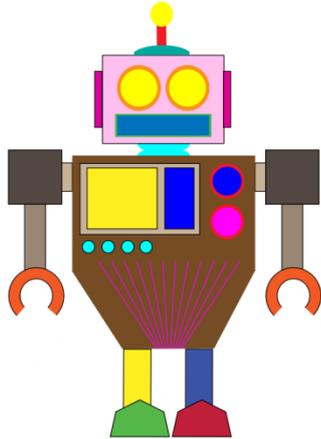
Inhalt des Vortrags

Motivation

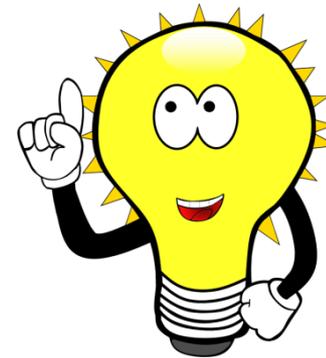
Worauf kommt es bei der Suche nach verborgenen Mustern an?

Inhalt des Vortrags

Data Science im Unternehmen



Verbesserung/ Automatisierung
von Geschäftsprozessen



Neue Erkenntnisse/
neue Zusammenhänge

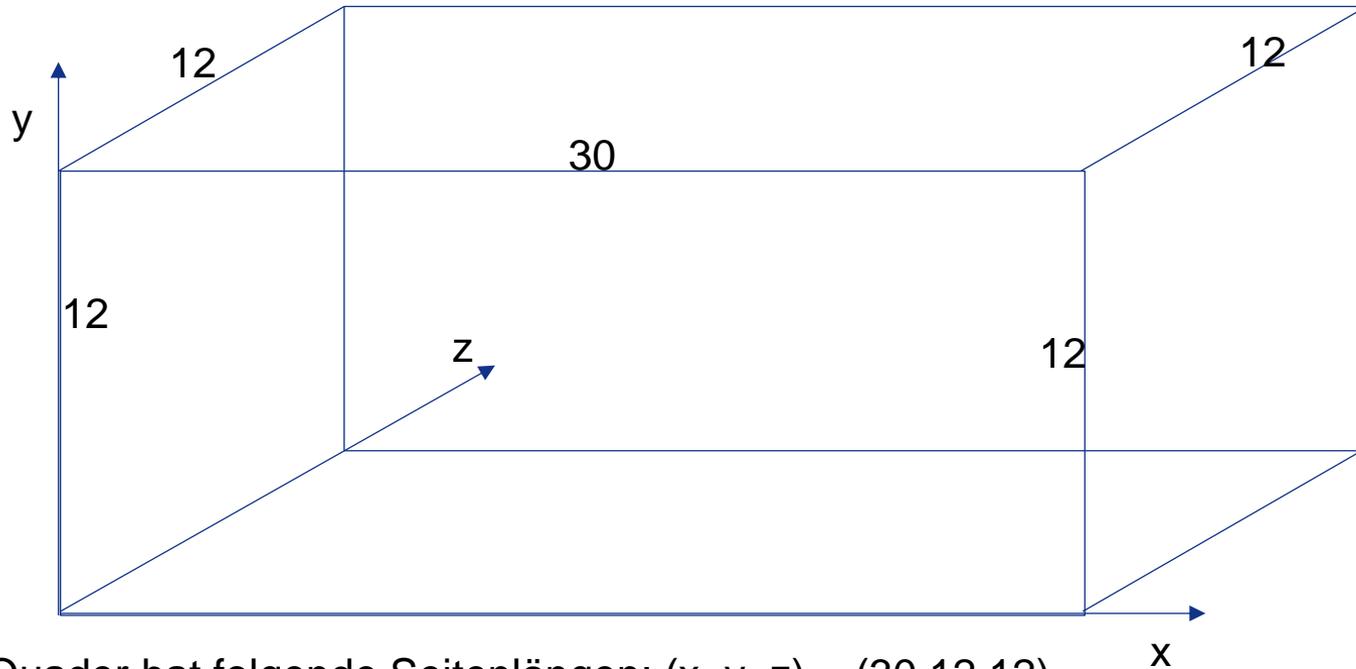


Starten wir doch mit einem Rätsel...



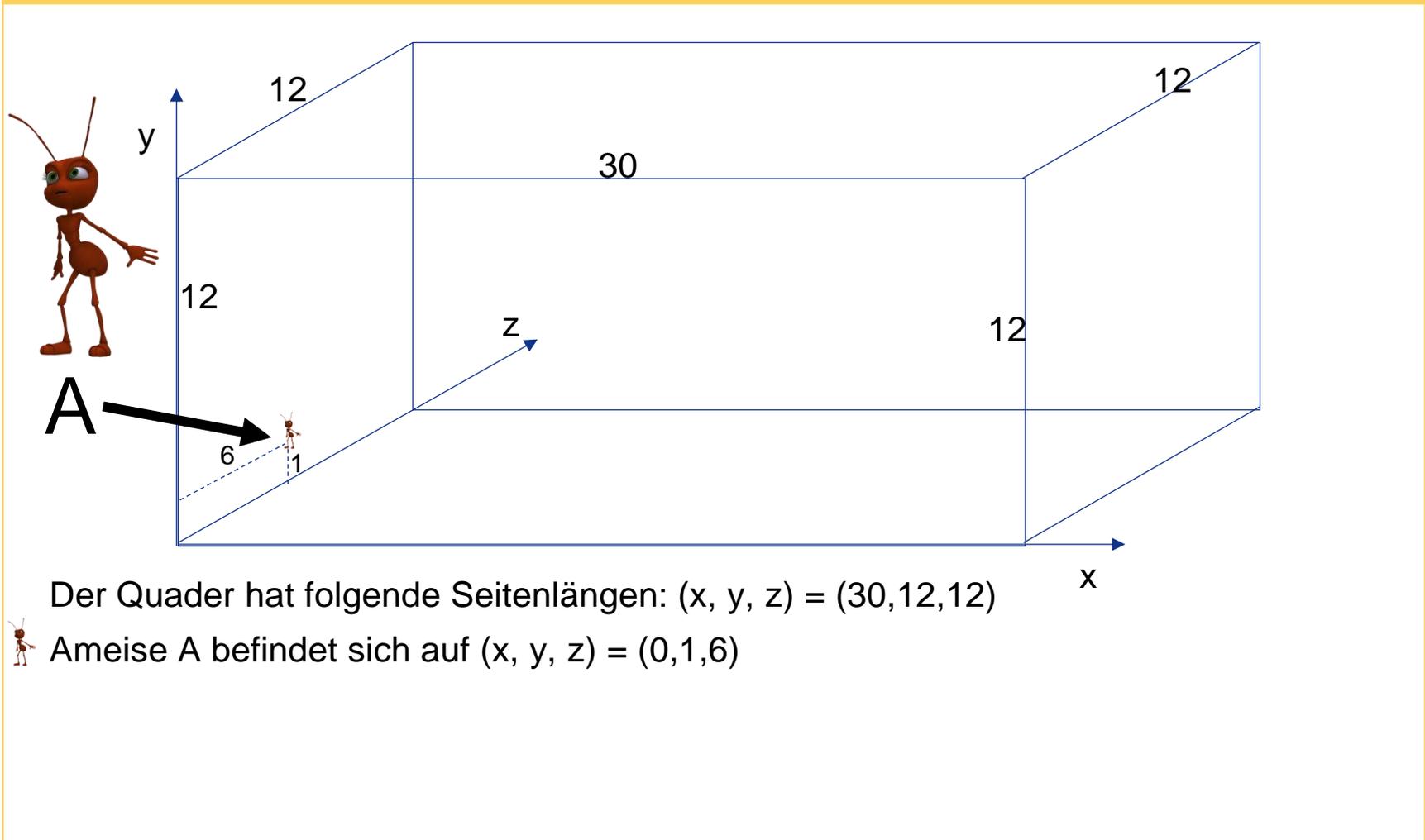


Starten wir doch mit einem Rätsel...



Der Quader hat folgende Seitenlängen: $(x, y, z) = (30, 12, 12)$

Starten wir doch mit einem Rätsel...



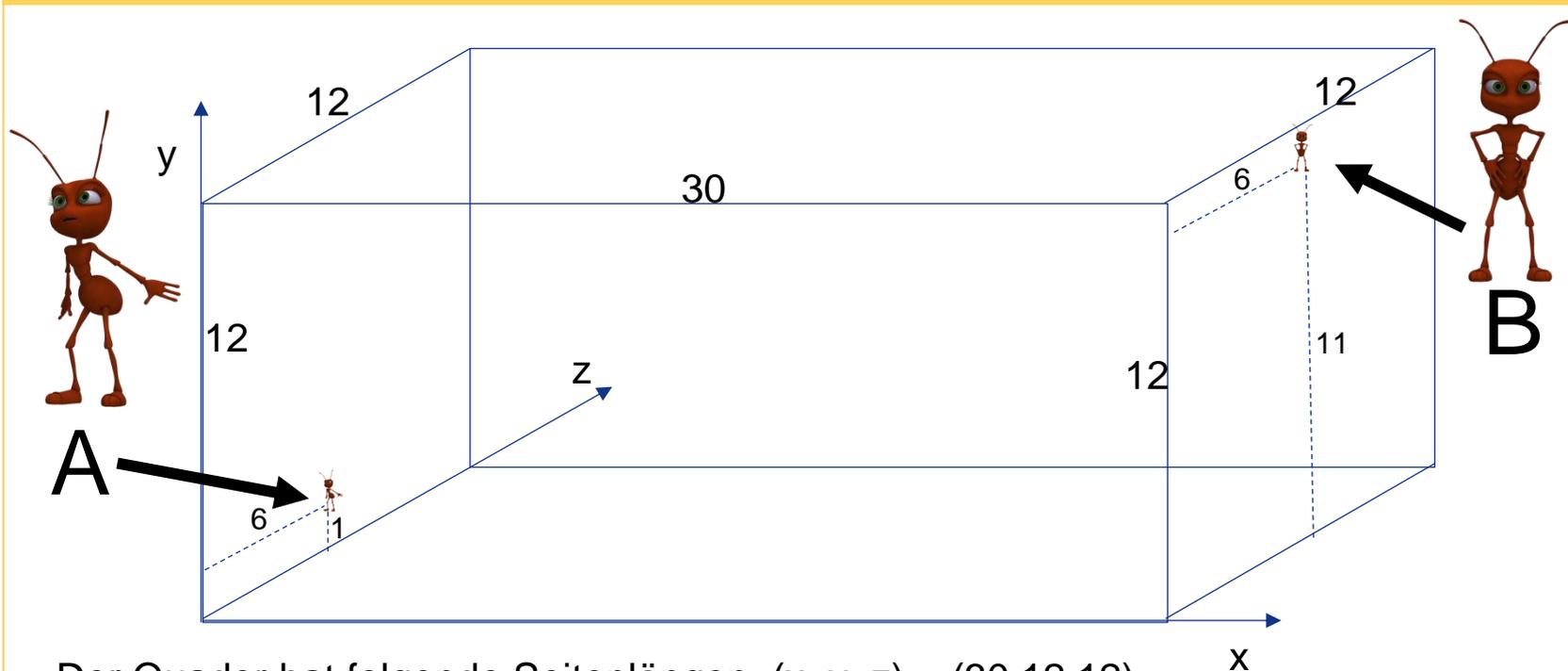
Starten wir doch mit einem Rätsel...

Der Quader hat folgende Seitenlängen: $(x, y, z) = (30, 12, 12)$

Ameise A befindet sich auf $(x, y, z) = (0, 1, 6)$

Ameise B befindet sich auf $(x, y, z) = (30, 11, 6)$

Starten wir doch mit einem Rätsel...



Der Quader hat folgende Seitenlängen: $(x, y, z) = (30, 12, 12)$

 Ameise A befindet sich auf $(x, y, z) = (0, 1, 6)$

 Ameise B befindet sich auf $(x, y, z) = (30, 11, 6)$

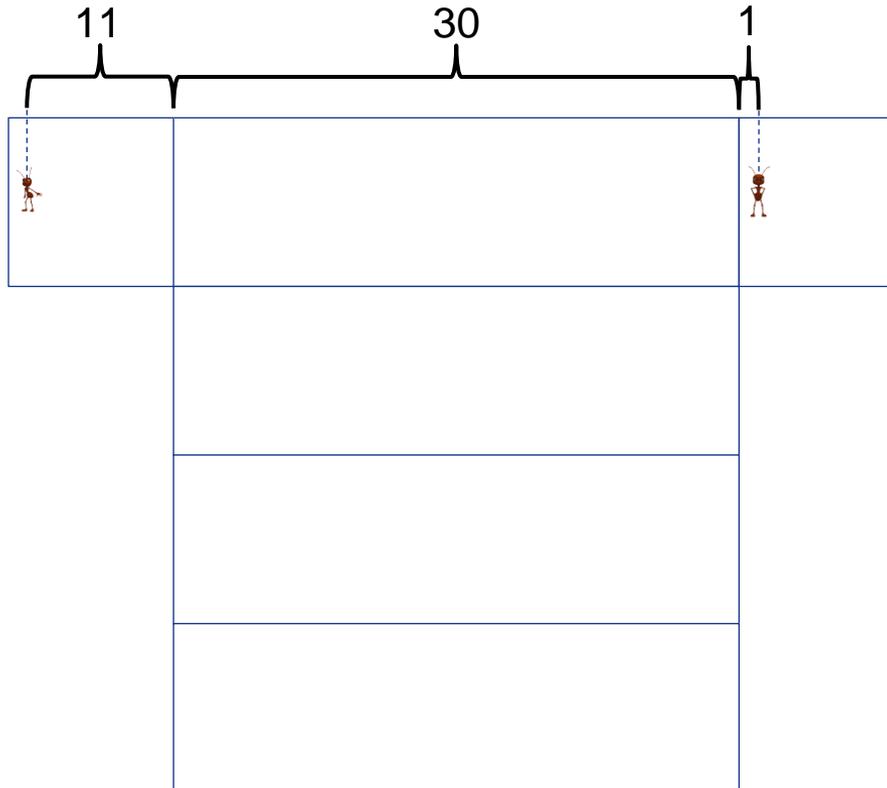
Quiz: Was ist der kürzeste Weg für Ameise A, um zu Ameise B zu kommen?

(Ameise B bewegt sich nicht und für den Weg ist nur das Bewegen auf der Oberfläche möglich)



Starten wir doch mit einem Rätsel...

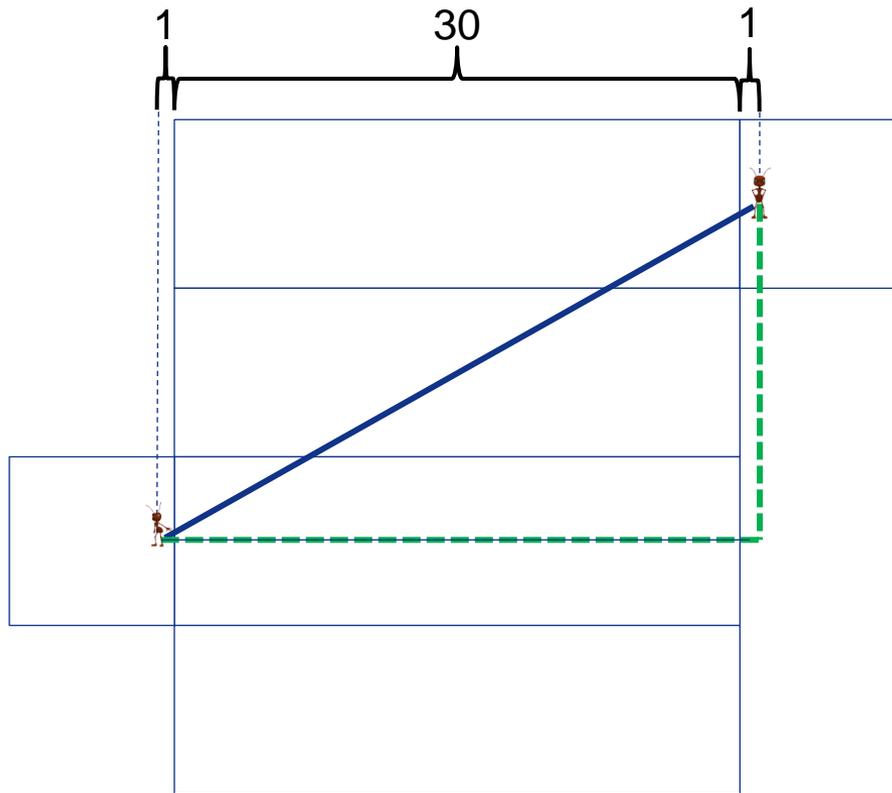
Nach Aufklappen des Quaders...





Starten wir doch mit einem Rätsel...

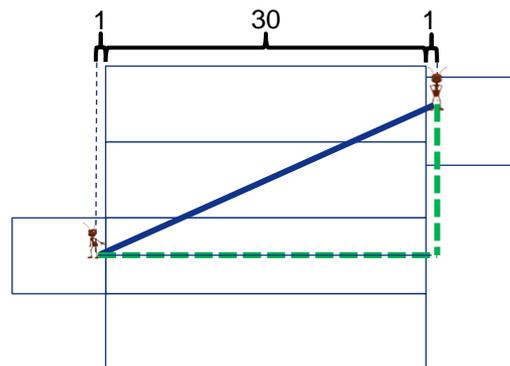
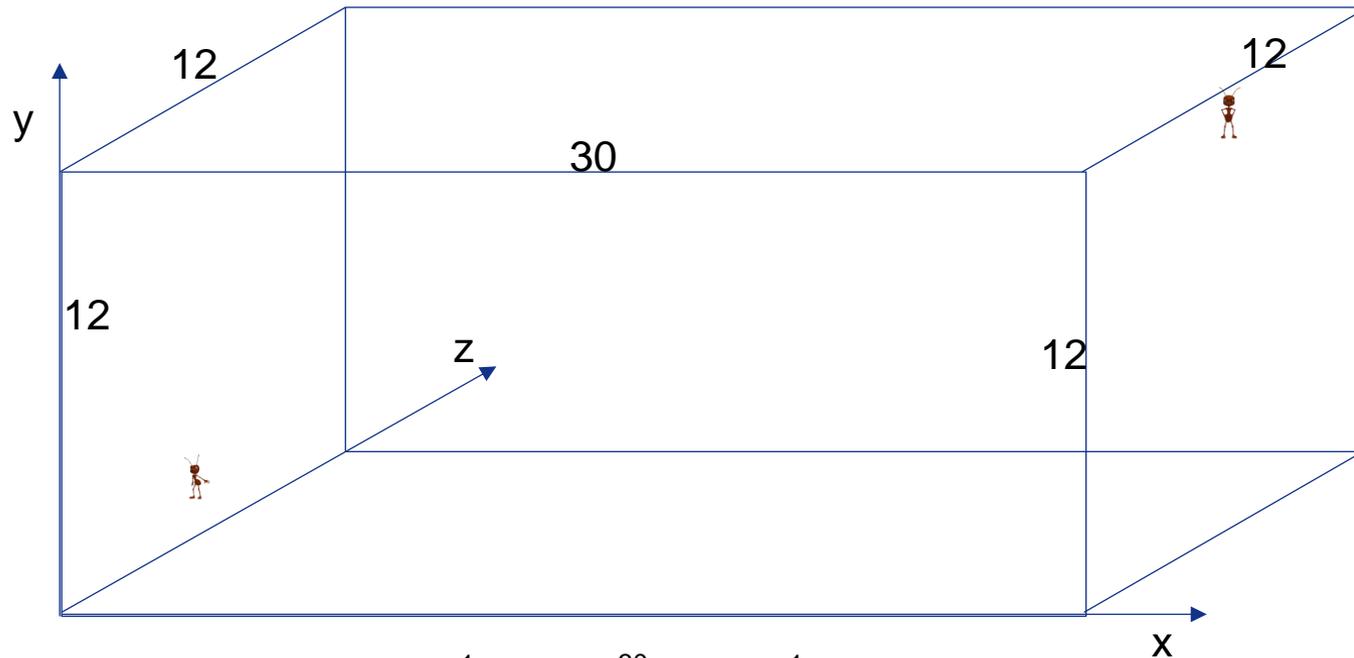
Antwort: Der kürzeste Weg ist nur 40 Einheiten lang!



Repräsentation!

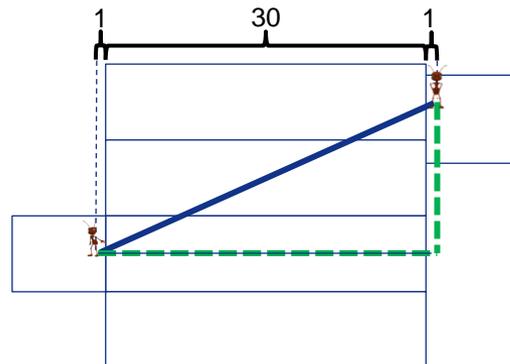
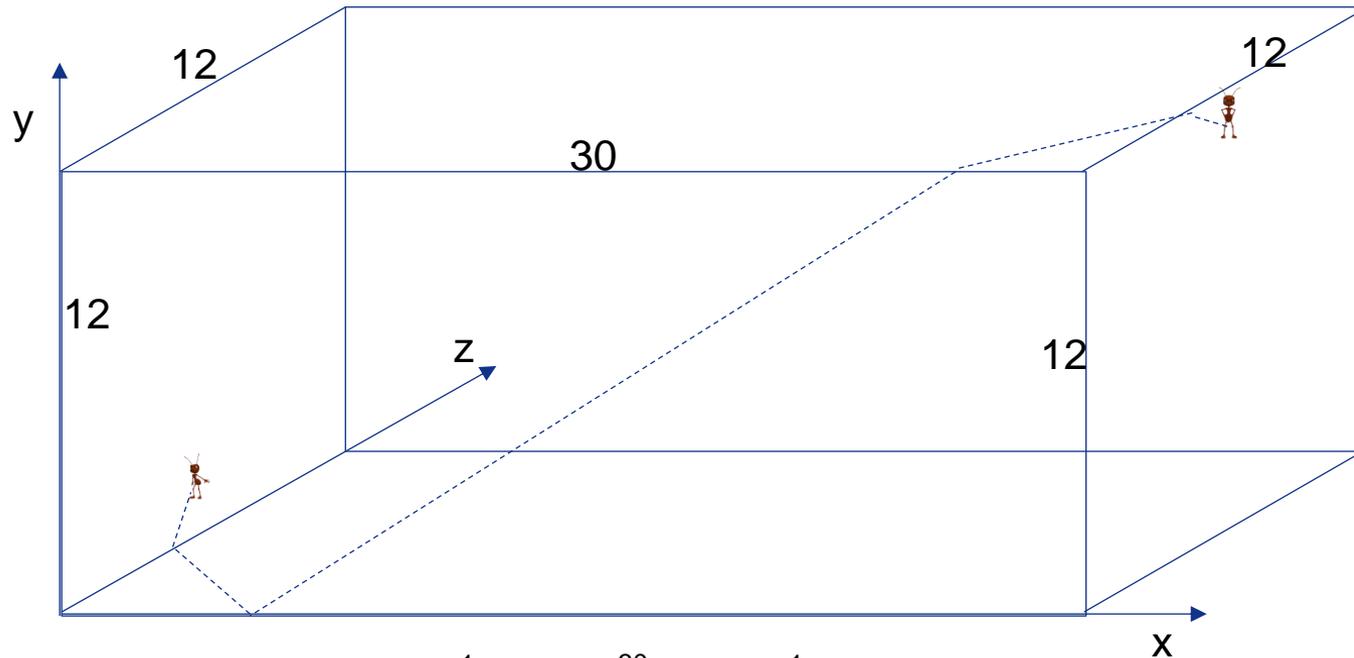


Starten wir doch mit einem Rätsel...





Starten wir doch mit einem Rätsel...





Starten wir doch mit einem Rätsel...

Lessons Learned #1

Datenvorverarbeitung:

Transformiere Daten in eine für die Fragestellung passende Repräsentation



Ach ja, da war noch was...

Der Startpunkt einer Analyse

Anforderungen an den Data Scientisten
sind **niemals** klar formuliert

Ach ja, da war noch was...

Lessons Learned #2

Nur mit guten kommunikative Fähigkeiten können Sie Antworten auf Fragen finden (die man Ihnen so nicht gestellt hat)





Wie hoch ist die Vertragsdichte in Branche x?

Vertragsdichte in verschiedenen Branchen

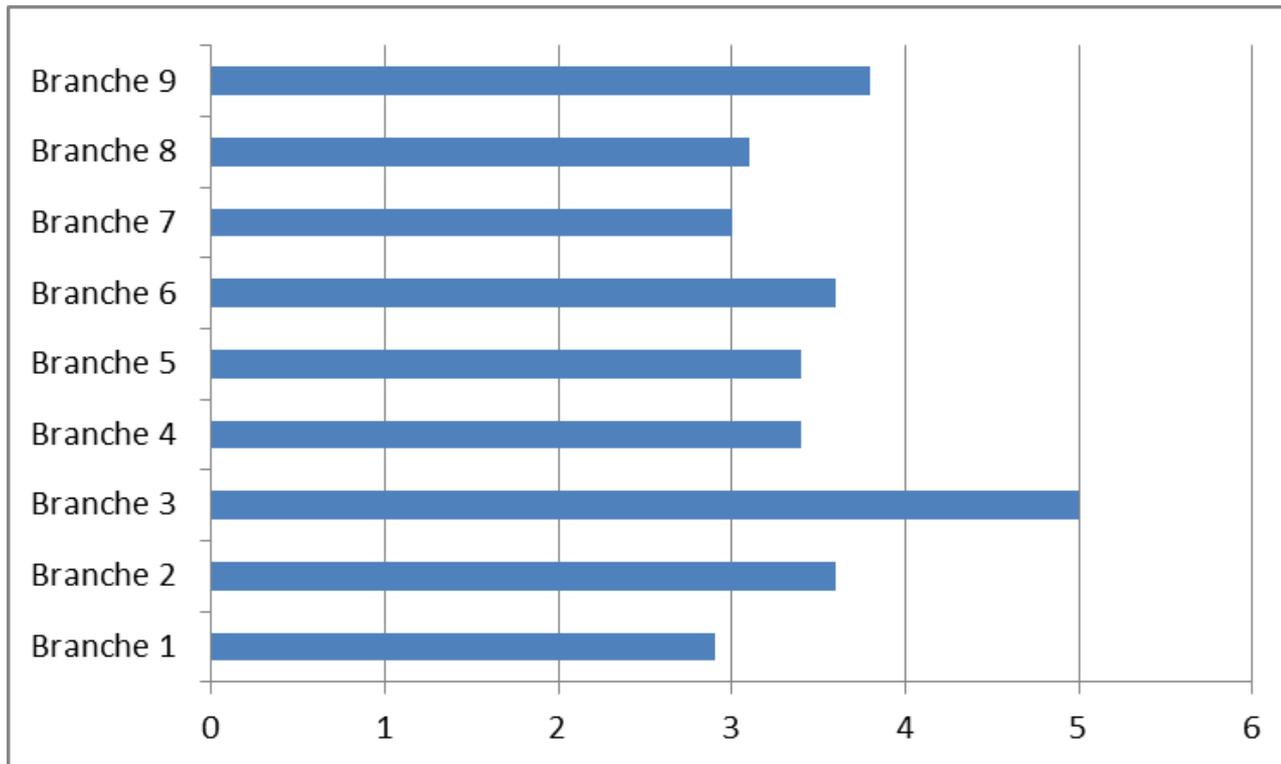
Wie hoch ist die Vertragsdichte in unserer Branche im Vergleich zum Gesamtbestand?



Wie hoch ist die Vertragsdichte in Branche x?

Vertragsdichte in verschiedenen Branchen

illustrativ



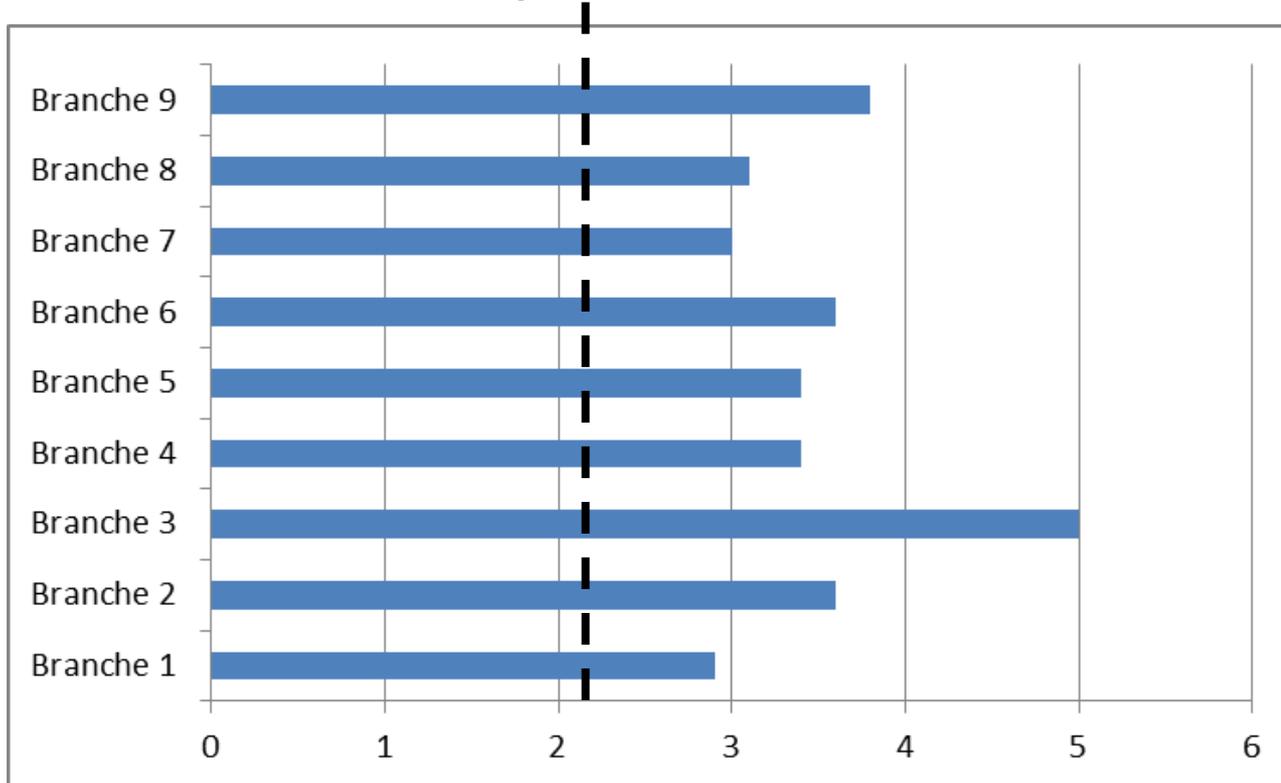


Wie hoch ist die Vertragsdichte in Branche x?

Vertragsdichte in verschiedenen Branchen

illustrativ

Mittelwert Vertragsdichte über Bestand





Wie hoch ist die Vertragsdichte in Branche x?

Das Bäumler'sche Paradoxon

illustrativ

	Sach	Leben	Kranken
Kunde 1	1	0	1
Kunde 2	0	1	0
Kunde 3	0	0	1
Kunde 4	1	1	1



Wie hoch ist die Vertragsdichte in Branche x?

Das Bäumler'sche Paradoxon

illustrativ

	Sach	Leben	Kranken	# Verträge
Kunde 1	1	0	1	2
Kunde 2	0	1	0	1
Kunde 3	0	0	1	1
Kunde 4	1	1	1	3
# Kunden	2	2	3	

Anzahl Verträge: 7

Anzahl Kunden: 4

Vertragsdichte (gesamt): $7/4 = 1,75$



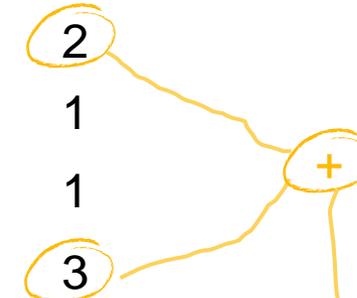
Wie hoch ist die Vertragsdichte in Branche x?

Das Bäumler'sche Paradoxon

illustrativ

	Sach	Leben	Kranken
Kunde 1	1	0	1
Kunde 2	0	1	0
Kunde 3	0	0	1
Kunde 4	1	1	1
# Kunden	2	2	3

Verträge



Anzahl Sachkunden: 2
Anzahl Lebenskunden: 2
Anzahl Krankenkunden: 3

Anzahl Verträge: 7
Anzahl Kunden: 4
Vertragsdichte (gesamt): $7/4 = 1,75$

Vertragsdichte Sachkunden: $5/2 = 2,5$
Vertragsdichte Lebenskunden: $4/2 = 2$
Vertragsdichte Krankenkunden: $6/3 = 2$



Wie hoch ist die Vertragsdichte in Branche x?

Lessons Learned #3

Wähle eine geeignete Referenzgruppe



Nichts als Korrelationen...

Korrelation ist nicht gleich Kausalität

Korrelation \neq Kausalität

Korrelation \neq Kausalität

Korrelation \neq Kausalität

Korrelation \neq Kausalität



Nichts als Korrelationen...

Lessons Learned #4

Nur mit fachlichem Wissen können Sie
Kausalität von Korrelationen unterscheiden

Welche Produkte werden gemeinsam gekauft?

Market basket analysis



# customers	fruit	beer	candy	magazines	...
6.388.860	1	0	0	0	
898.973	1	0	1	0	
4.231.452	0	1	0	0	
5.123.433	0	1	1	1	
...	

Welche Produkte werden gemeinsam gekauft?

Sortieren nach Häufigkeit...



# customers	fruit	beer	candy	magazines	...
6.567.680	1	1	0	0	
6.549.840	1	1	1	0	
6.488.320	1	0	1	0	
6.388.860	1	0	0	0	
...	



Welche Produkte werden gemeinsam gekauft?

... oder der Erstellen einer Pivottabelle...

Zeilenbeschriftungen		Summe von _customers
=0		20089368
=0		9965592
=0	0	4823352
=1		5142240
=0	0	1267040
=0	0	632864
=0	0	315776
=0	0	157232
=0	0	77960
=0	0	77960
=0	0	46840
=0	0	23020
=0	0	23020
=0	0	23020
=0	0	8482
=1		8482
=0	0	8482
=0	0	1213
=0	0	1213
=0	0	256
=0	0	256
=1		957
=0	0	957
=1		7269
=0	0	7269
=0	0	3252
=0	0	3252
=1		4017
=0	0	4017
=1		14538
=1		14538
=0	0	14538
=0	0	4209
=0	0	4209
=0	0	1722



Welche Produkte werden gemeinsam gekauft?

... oder eine Warenkorbanalyse gibt häufig nur begrenzt Erkenntnisse!



```
107 (b.wat) -> (f.wat) 0.226058 0.2921301 0.996022
108 (f.fil) -> (f.fil) 0.276782 0.5461100 1.0002072
109 (b.wat) -> (f.wat) 0.276782 0.5144172 1.0002072
110 (f.veg) -> (f.veg) 0.3124506 0.6165395 0.9872194
111 (b.wat) -> (f.fru) 0.3120869 0.6152119 1.0000144
112 (f.fru) -> (b.wat) 0.3120869 0.5162953 1.0069344
113 (b.jul) -> (f.meat) 0.2635273 0.5186947 0.9897580
114 (f.meat) -> (b.jul) 0.2635273 0.5187323 0.9897580
115 (b.jul) -> (f.fil) 0.2811916 0.5449267 1.0070285
116 (f.fil) -> (b.jul) 0.2811916 0.5196080 1.0070285
117 (b.jul) -> (f.veg) 0.3157756 0.6110476 0.9798669
118 (f.veg) -> (b.jul) 0.3157756 0.5056283 0.9798669
119 (b.jul) -> (f.fru) 0.3165203 0.6133988 1.0029593
120 (f.fru) -> (b.jul) 0.3165203 0.5175444 1.0029593
121 (f.meat) -> (f.fil) 0.2888593 0.5598271 1.0345563
122 (f.fil) -> (f.meat) 0.2888593 0.5338997 1.0345563
123 (f.meat) -> (f.veg) 0.3048155 0.5967513 0.9459268
124 (f.veg) -> (f.meat) 0.3048155 0.4880797 0.9459268
125 (f.meat) -> (f.fru) 0.3241218 0.6281681 1.0271218
126 (f.fru) -> (f.meat) 0.3241218 0.5299737 1.0271218
127 (f.fil) -> (f.veg) 0.3065035 0.5664661 0.9069605
128 (f.veg) -> (f.fil) 0.3065035 0.4907815 0.9069605
129 (f.fil) -> (f.fru) 0.3417171 0.6114986 1.0325545
130 (f.fru) -> (f.fil) 0.3417171 0.5587439 1.0325545
131 (f.veg) -> (f.fru) 0.3299639 0.5283470 0.8639036
132 (f.fru) -> (f.veg) 0.3299639 0.5395261 0.8639036
133 (f.can,F.can) -> (f.veg) 0.1578878 0.6299226 1.0022220
134 (f.can,F.veg) -> (f.can) 0.1578878 0.5025113 0.9993173
135 (f.can,F.veg) -> (f.can) 0.1578878 0.5024205 0.9994376
136 (f.can,F.can) -> (f.can) 0.1546667 0.6131923 1.0001884
137 (f.can,F.fru) -> (f.can) 0.1546667 0.5026169 0.9995273
138 (f.can,F.fru) -> (f.can) 0.1546667 0.5025938 0.9995914
139 (f.egg,F.can) -> (f.veg) 0.1579488 0.6250822 1.0020511
140 (f.can,F.veg) -> (f.egg) 0.1579488 0.5026801 0.9990752
141 (f.egg,F.veg) -> (f.can) 0.1579488 0.5024257 0.9994360
142 (f.egg,F.can) -> (f.fru) 0.1547359 0.6131035 1.0024896
143 (f.can,F.fru) -> (f.egg) 0.1547359 0.5028420 0.9993970
144 (f.egg,F.fru) -> (f.can) 0.1547359 0.5025227 0.9993891
145 (b.mil,F.can) -> (f.veg) 0.1574137 0.6236467 0.9985998
146 (f.can,F.veg) -> (b.mil) 0.1574137 0.5018025 0.9957151
147 (b.mil,F.veg) -> (f.can) 0.1574137 0.5020946 0.9987724
148 (b.mil,F.can) -> (f.fru) 0.1515116 0.6146844 1.0050746
149 (f.can,F.fru) -> (b.mil) 0.1515116 0.5041927 1.0015523
150 (b.mil,F.fru) -> (f.can) 0.1515116 0.5024959 0.9995757
151 (b.win,F.can) -> (f.veg) 0.1580469 0.6255017 1.0015701
152 (f.can,F.veg) -> (b.win) 0.1580469 0.5030177 0.9985910
153 (b.win,F.veg) -> (f.can) 0.1580469 0.5024241 0.9994327
154 (b.win,F.can) -> (f.fru) 0.1548745 0.6129462 1.0022324
155 (f.can,F.fru) -> (b.win) 0.1548745 0.5032223 0.9991369
156 (b.win,F.fru) -> (f.can) 0.1548745 0.5025004 0.9995846
157 (b.bee,F.can) -> (f.veg) 0.1578409 0.6242435 0.9967051
158 (f.can,F.veg) -> (b.bee) 0.1578409 0.5023020 0.9924553
159 (b.bee,F.veg) -> (f.can) 0.1578409 0.5020889 0.9987660
160 (b.bee,F.can) -> (f.fru) 0.1557057 0.6140432 1.0040262
161 (f.can,F.fru) -> (b.bee) 0.1557057 0.5059936 1.0000320
162 (b.bee,F.fru) -> (f.can) 0.1557057 0.5024870 0.9995580
163 (b.wat,F.can) -> (f.veg) 0.1569873 0.6125769 0.9888007
164 (f.can,F.veg) -> (b.wat) 0.1569873 0.4996452 0.9859191
165 (b.wat,F.veg) -> (f.can) 0.1569873 0.4924388 0.9951620
```

Ausgabe von R Studio - arules Paket



Welche Produkte werden gemeinsam gekauft?

Demo Item Explorer → <https://github.com/EE2dev/item-explorer>

Info

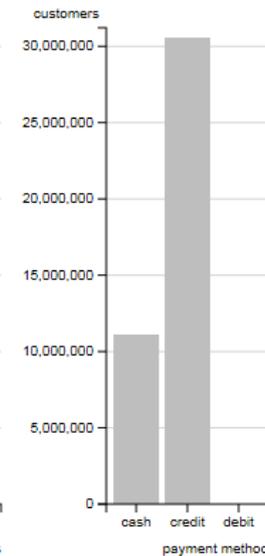
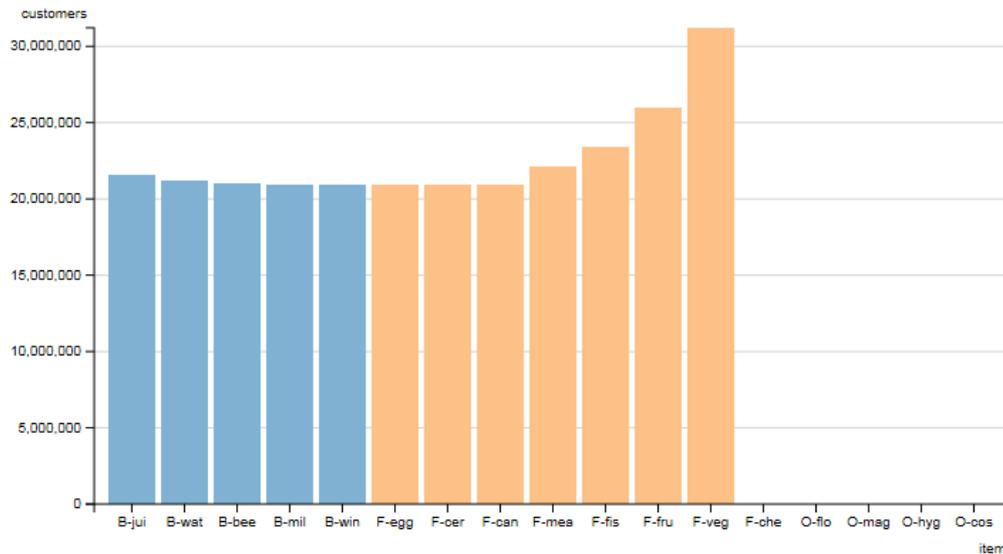
current filter:

frequency: 41,636,931 percent: 100%

Exploration

item 1	item 2	frequency	percent
F-veg	credit	24,307,588	58.4%
F-fru	credit	19,251,390	46.2%
F-fru	F-veg	18,205,890	43.7%

sort by frequency update axis





Welche Produkte werden gemeinsam gekauft?

Lesson Learned #5

Verknüpfe Fachwissen mit datengetriebenen Zusammenhängen:

Interaktive Visualisierungen



Zusammenfassung

Worauf kommt es bei der Suche nach verborgenen Mustern an?

- geeignete Datenrepräsentation
- gute kommunikative Fähigkeiten
- Wahl einer Referenzmenge
- fachliches Wissen zur Interpretation der Ergebnisse
- Verzahnung von fachlichem Wissen mit der Analyse