

Eksploracja danych

w identyfikacji praktyk antykonkurencyjnych

Łukasz Ziarko

Comper Fornalczyk i Wspólnicy sp. j.

Konferencja CARS i UOKiK

Warszawa 14.10.15

analiza praktyk wykluczających

analiza praktyk wykluczających
identyfikacja anomalii w danych

analiza praktyk wykluczających
identyfikacja anomalii w danych
segmentacja obserwacji

analiza praktyk wykluczających
identyfikacja anomalii w danych
segmentacja obserwacji
systematyzowanie danych o przetargach

analiza praktyk wykluczających

identyfikacja anomalii w danych

segmentacja obserwacji

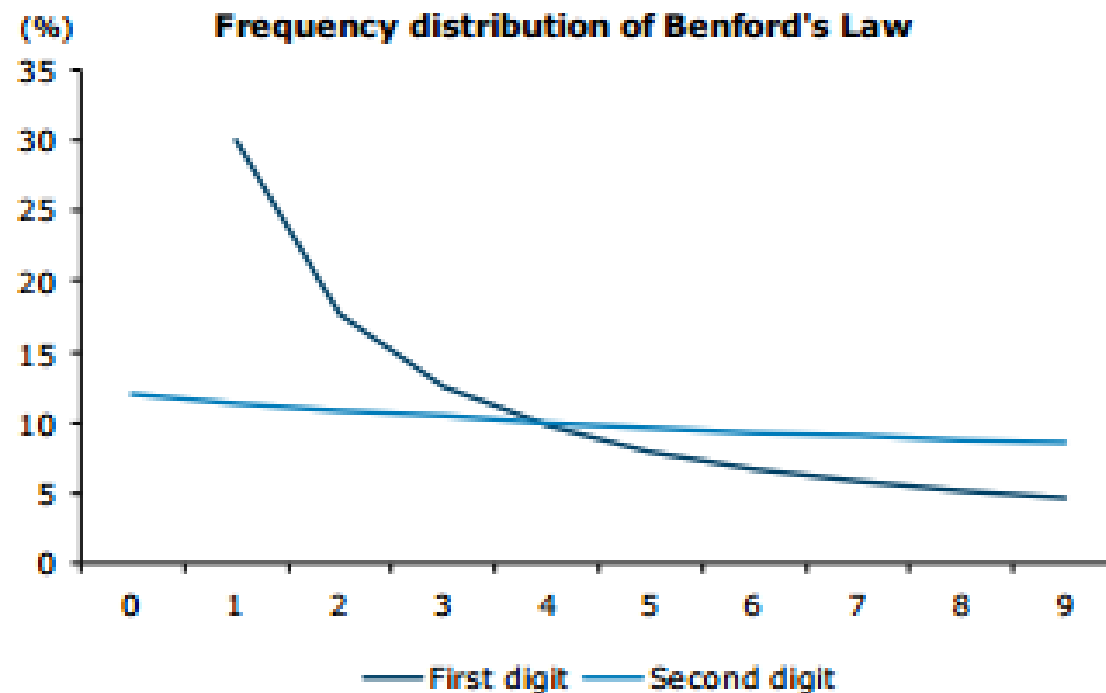
systematyzowanie danych o przetargach

definiowanie relacji biznesowych

$$P(\text{first_digit} = d) = \log_{10}(1 + d^{-1}), d = 1, 2, \dots, 9 \quad (1)$$

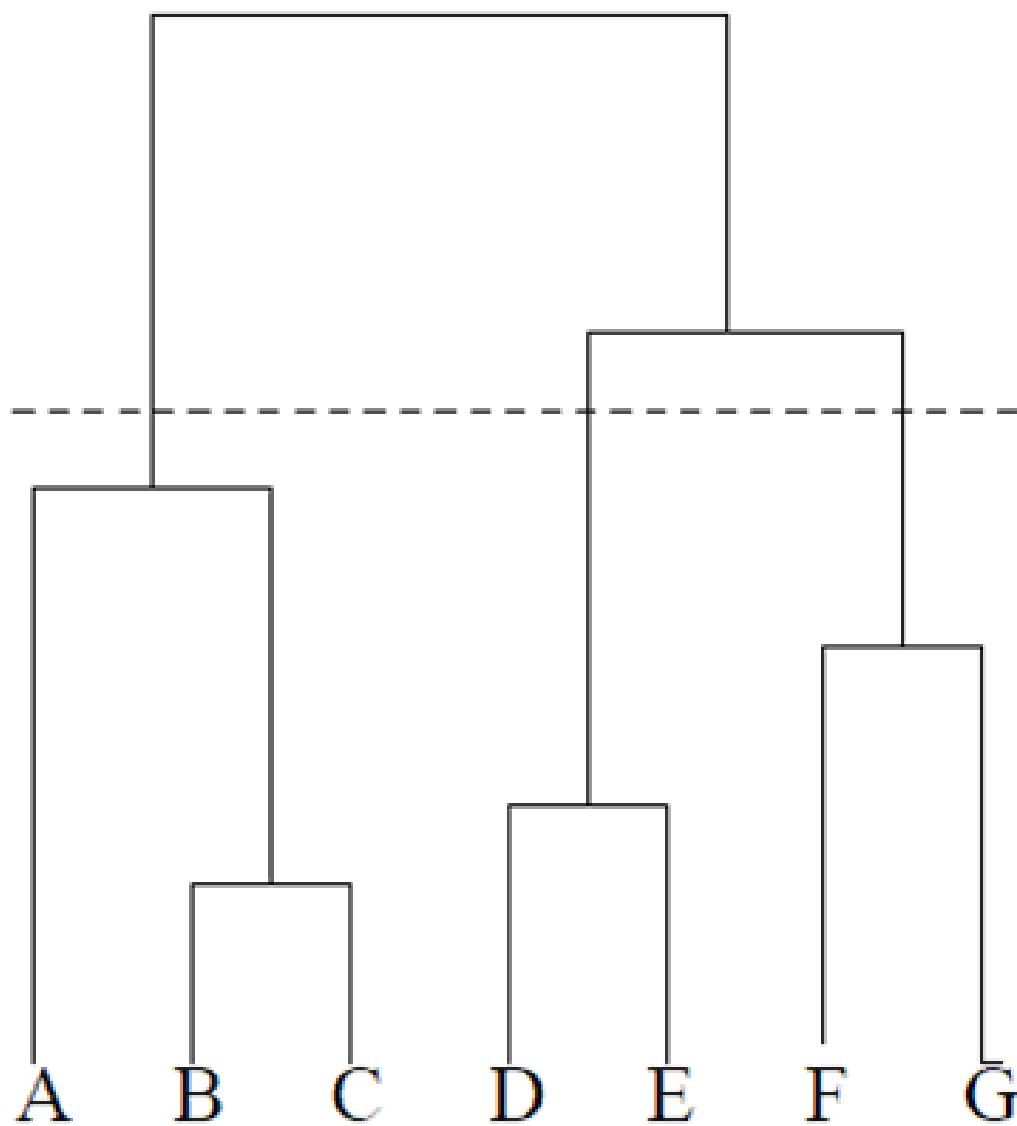
and

$$P(\text{second_digit} = d) = \sum_{k=1}^9 \log_{10}(1 + (10k + d)^{-1}), d = 0, 1, \dots, 9 \quad (2)$$

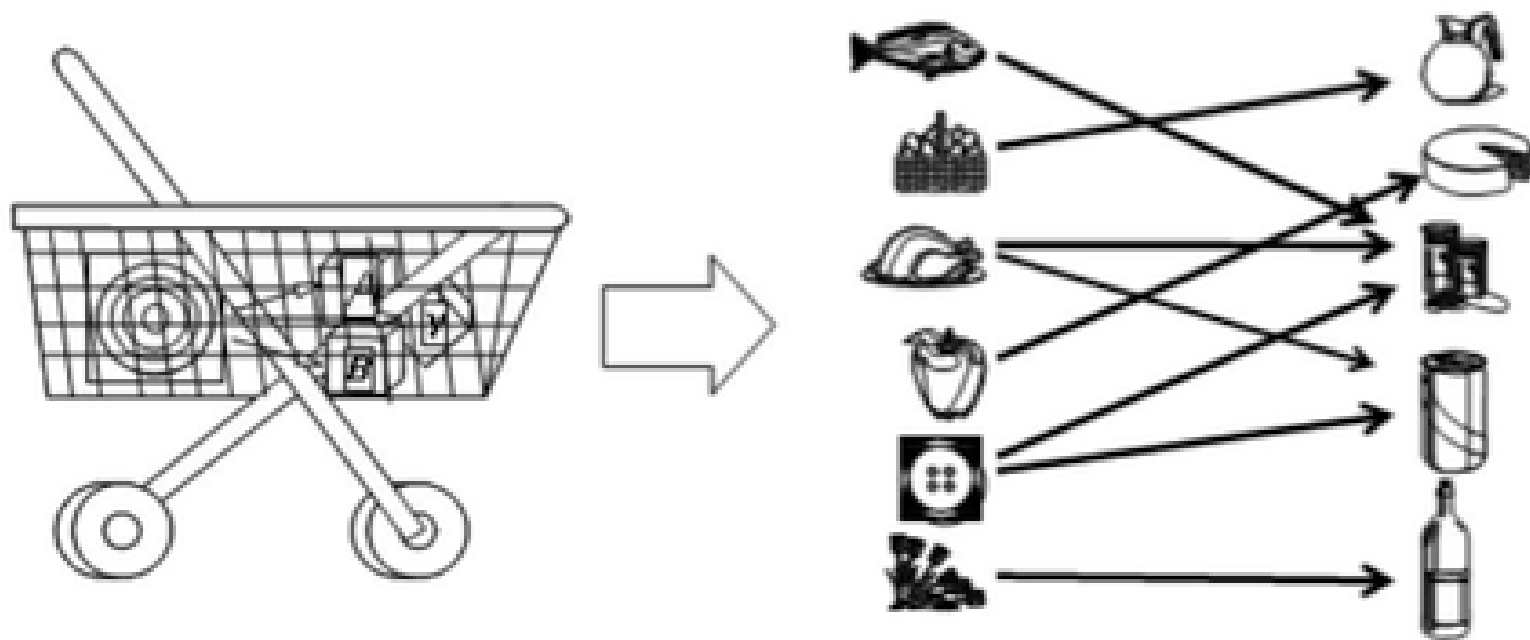


Source: ANZ

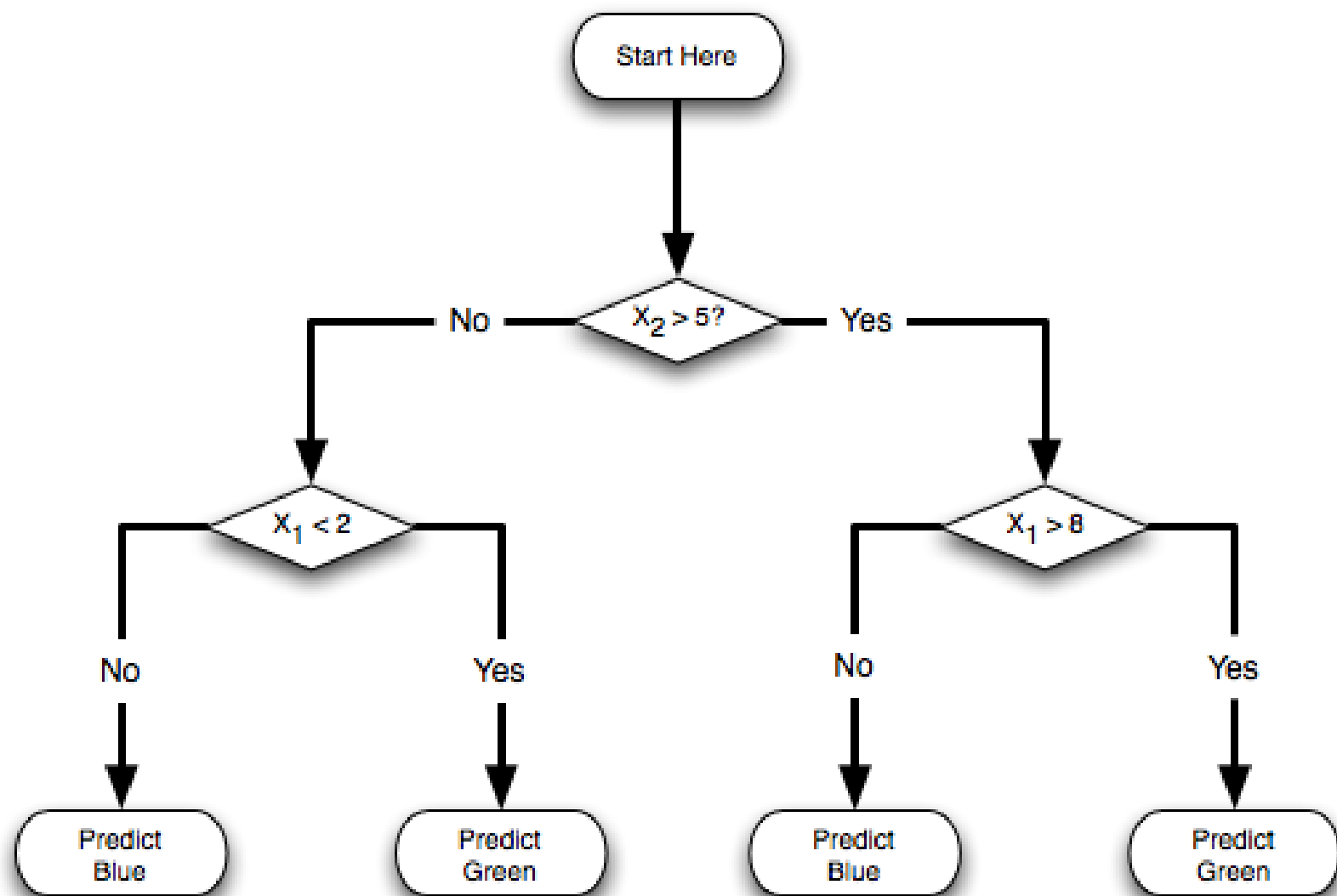
S
i
m
i
l
a
r
i
t
y



MARKET BASKET ANALYSIS



*98% of people who purchased items A and B
also purchased item C*



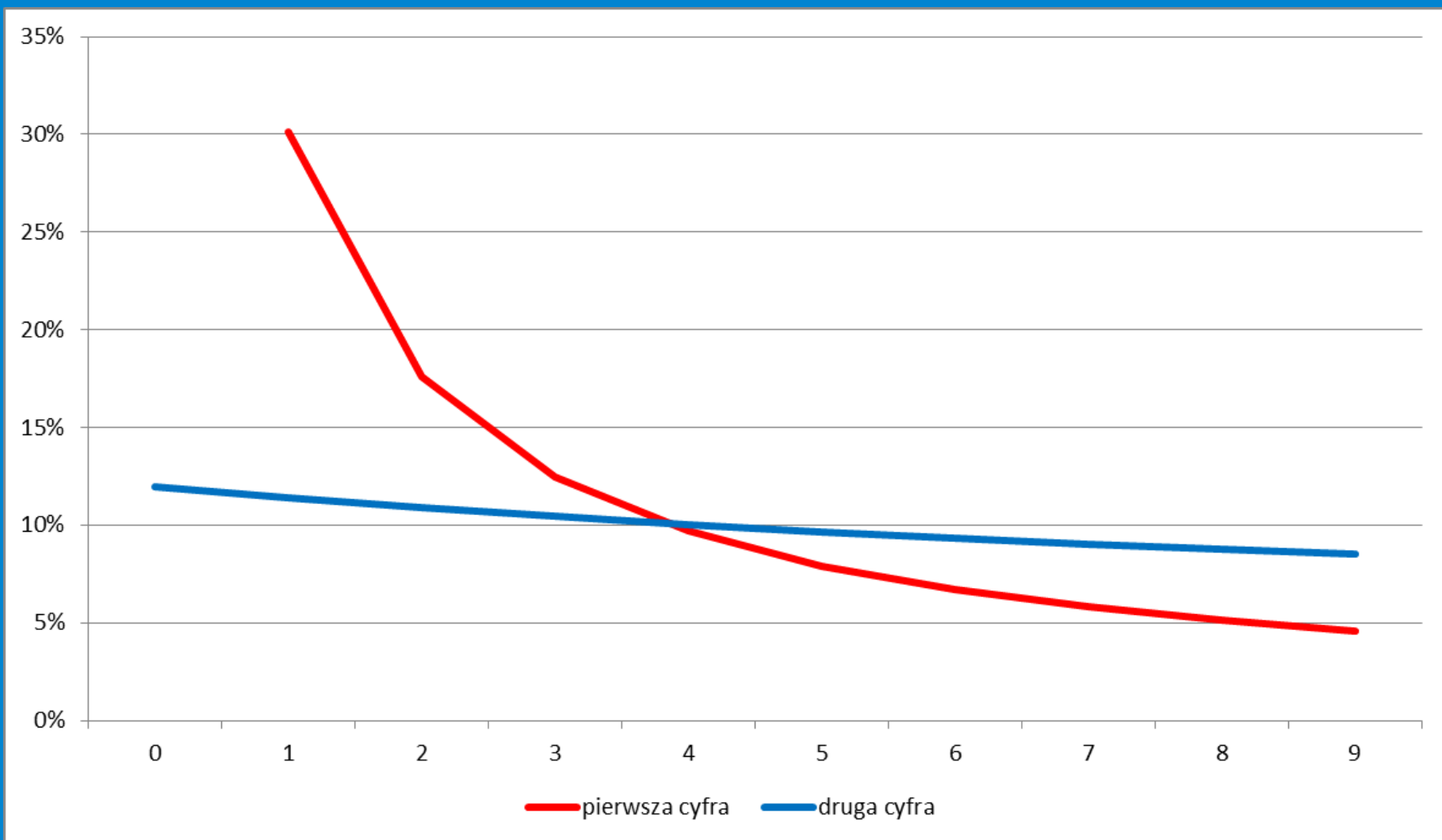


Figure II. 2nd Digit Fed. Funds Rate Empirical (1987-2005) & Benford Distributions

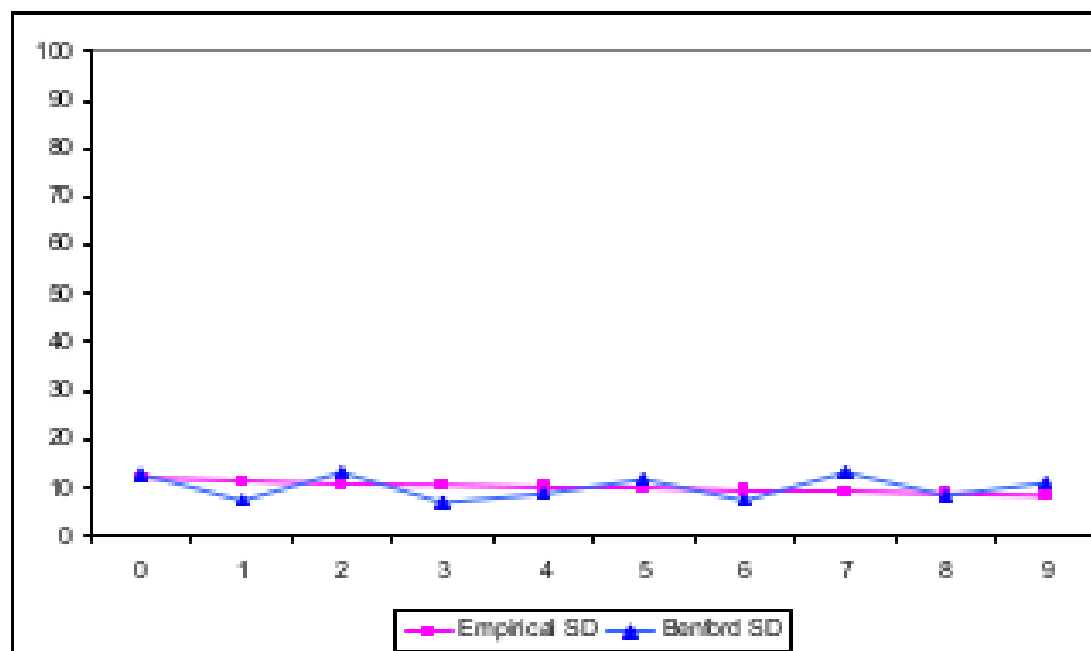


Figure III. Second Digit Labor Empirical (1987-2005) and Benford Distributions

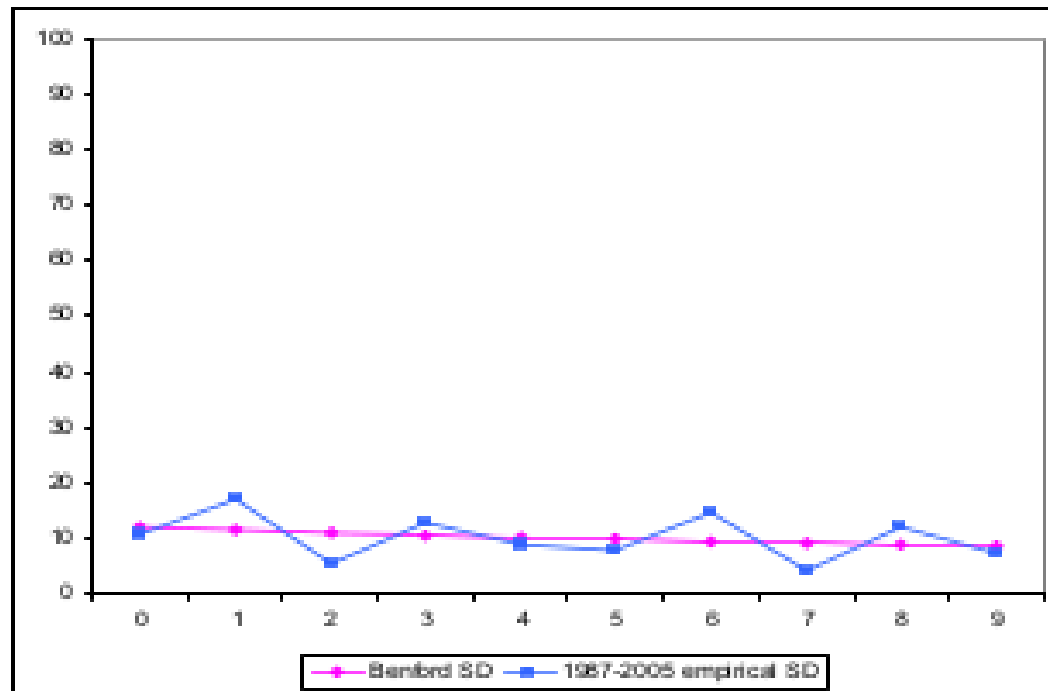
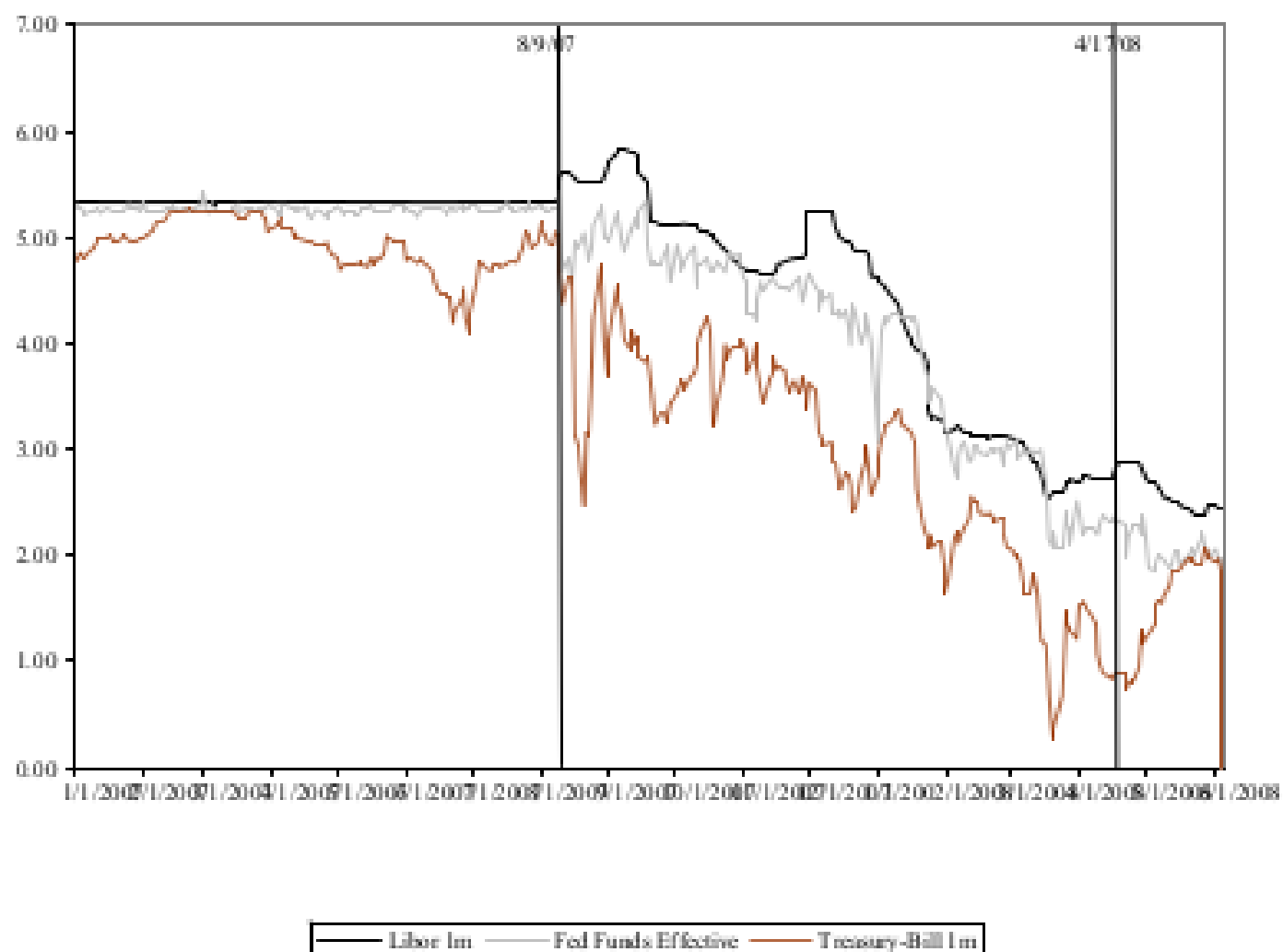


Figure IV. Libor 1m, Fed Funds Effective Rate and Treasury-Bill 1m



oferty kurtuazyjne

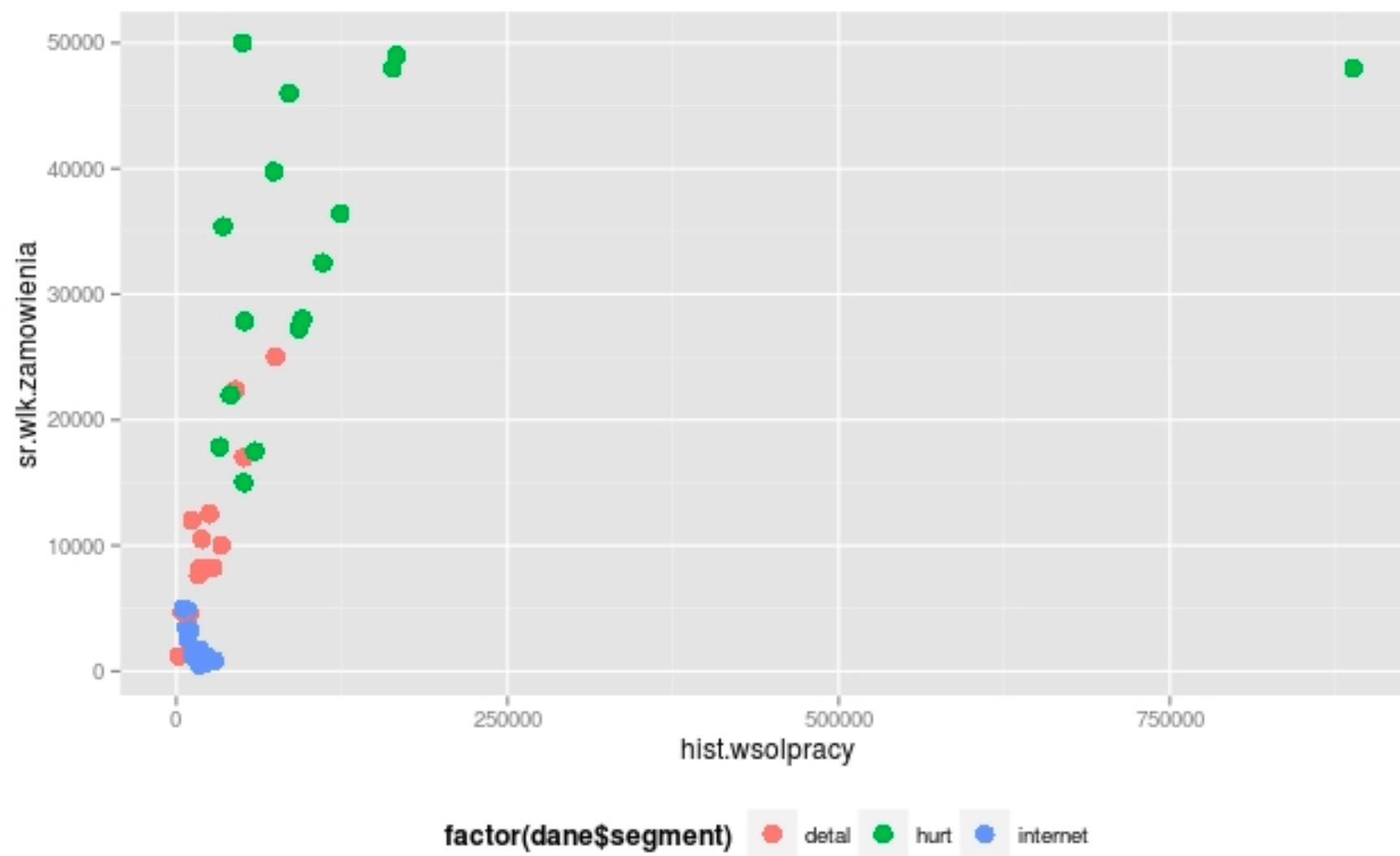
oferty kurtuazyjne
dyskryminacja cenowa

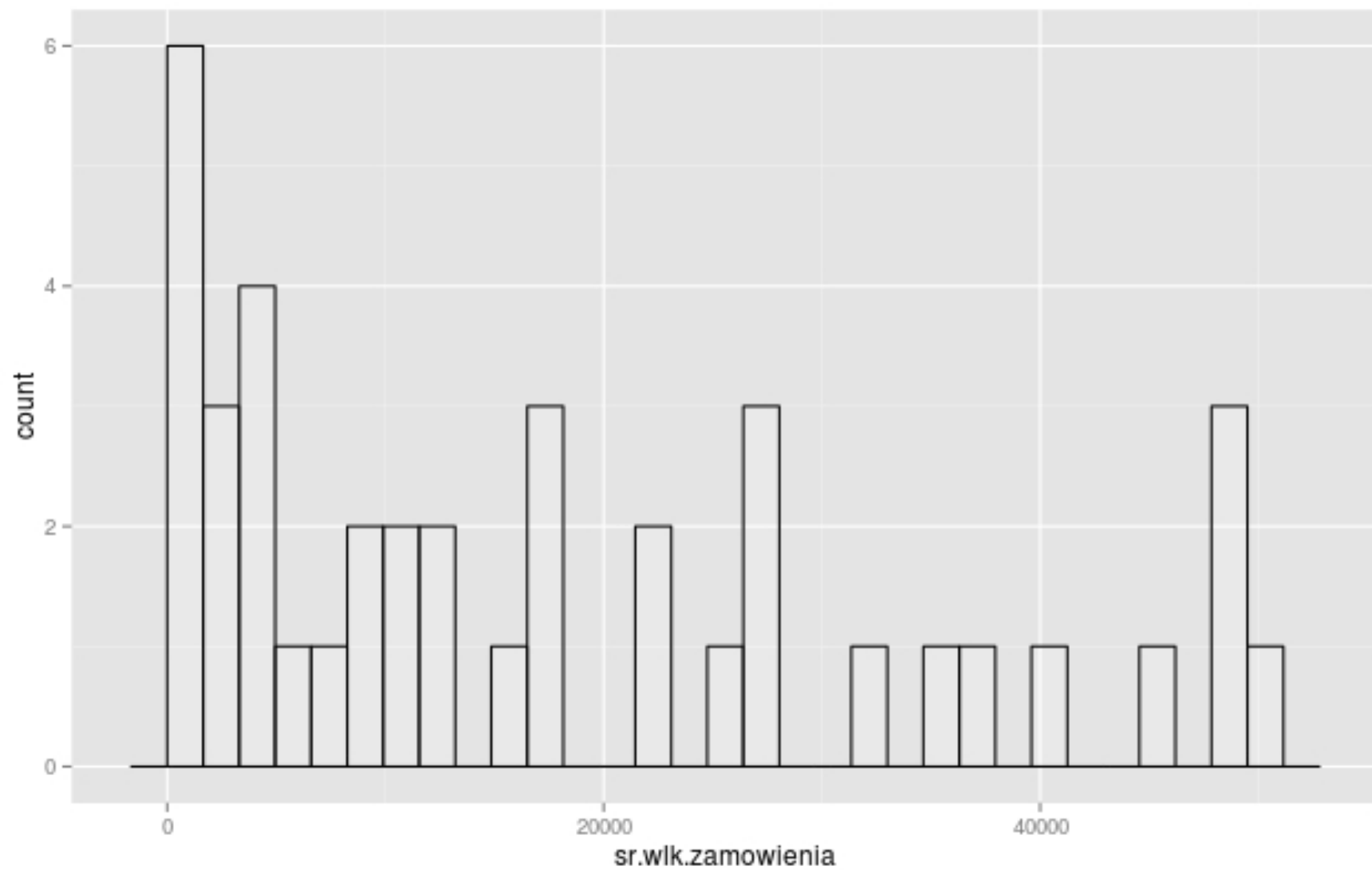
oferty kurtuazyjne

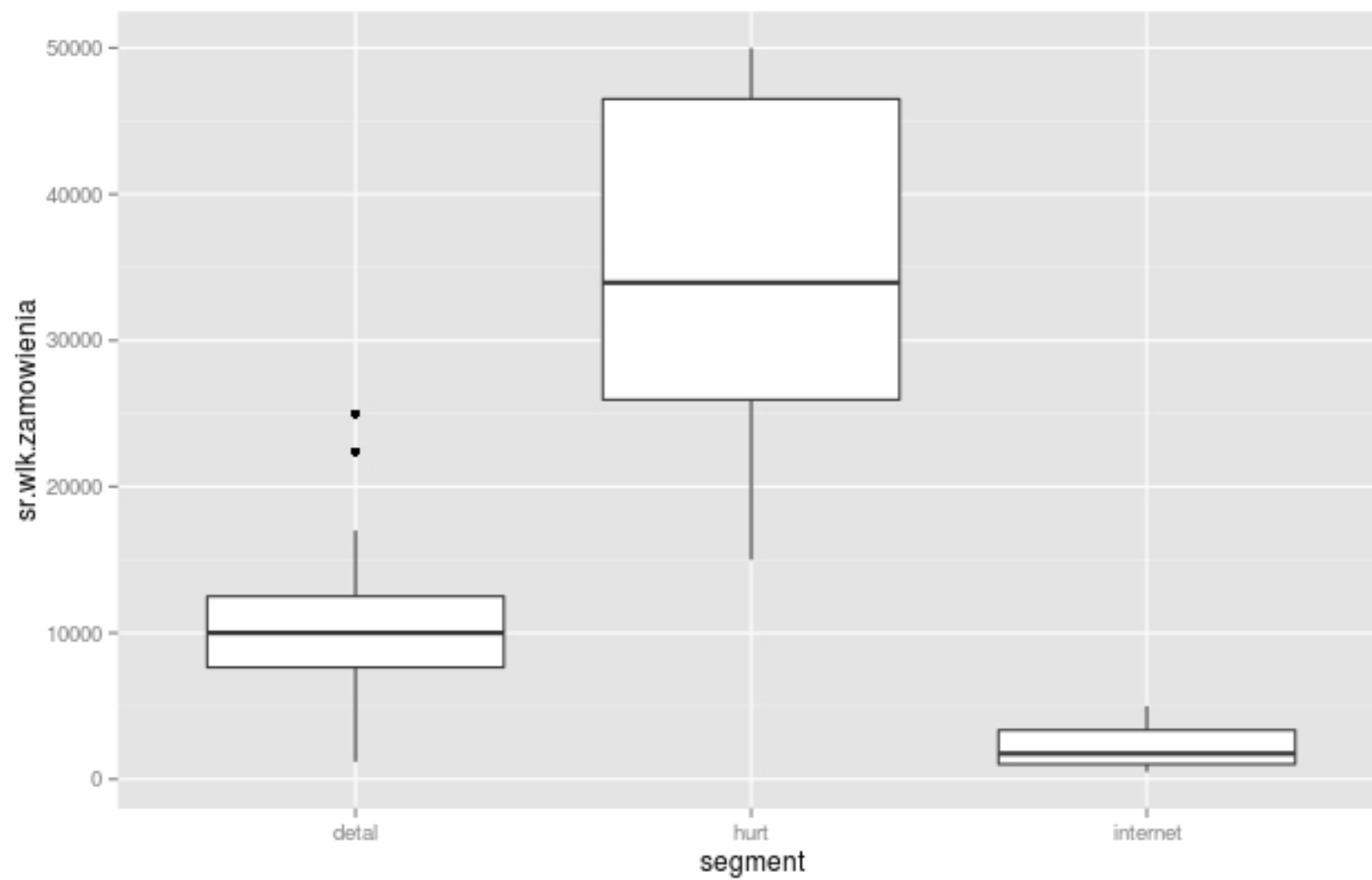
dyskryminacja cenowa

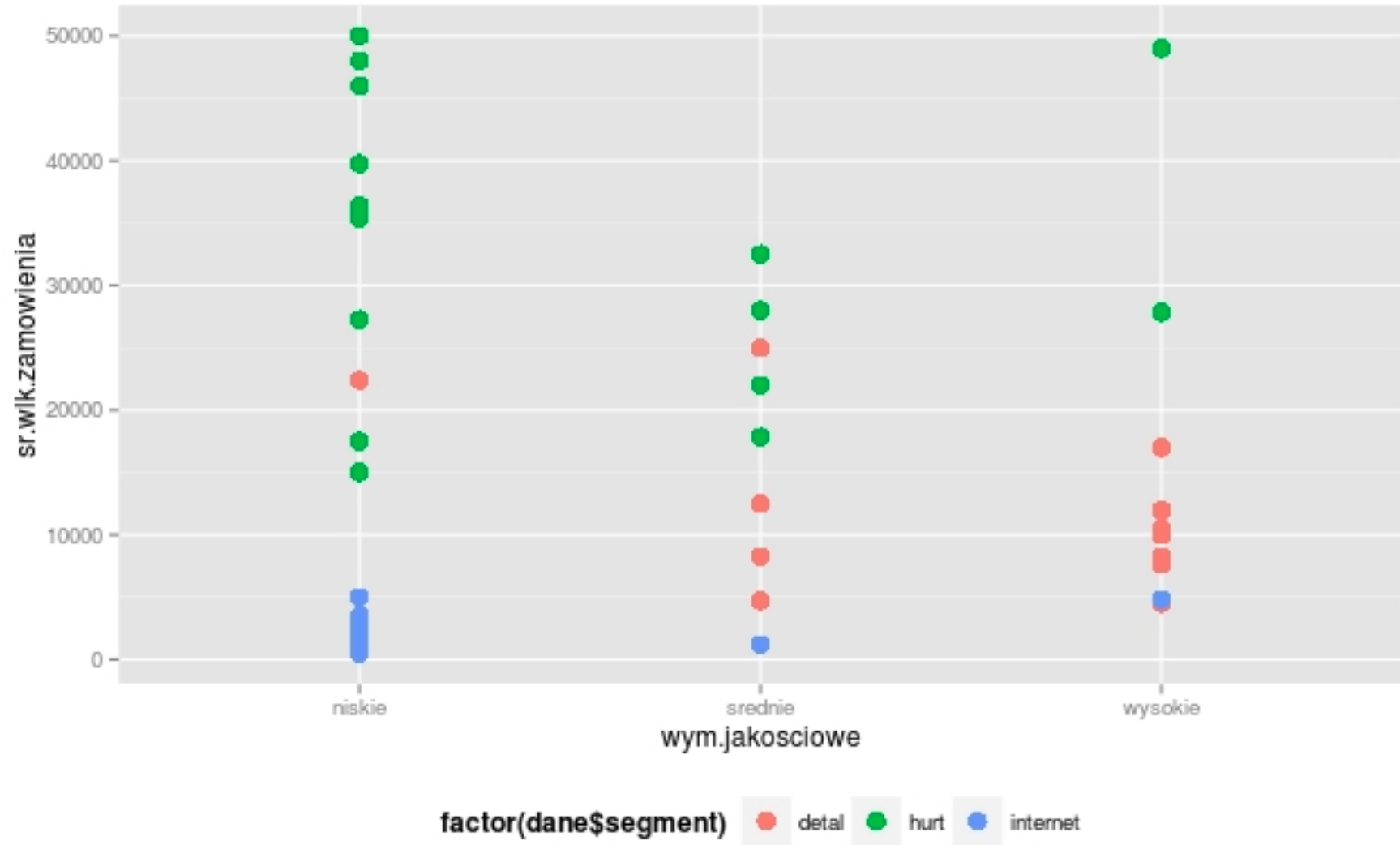
weryfikacja autentyczności danych

Podobieństwa

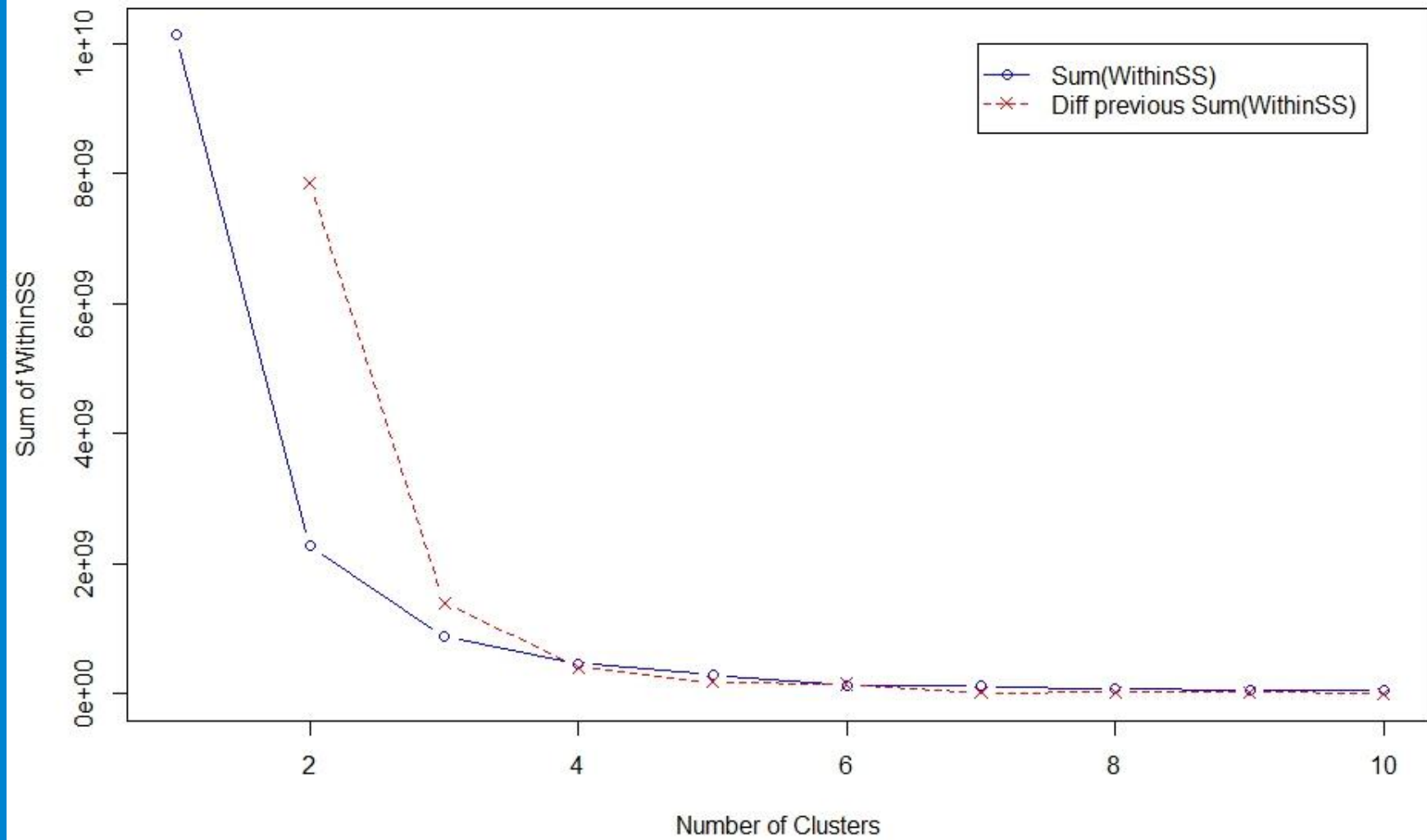


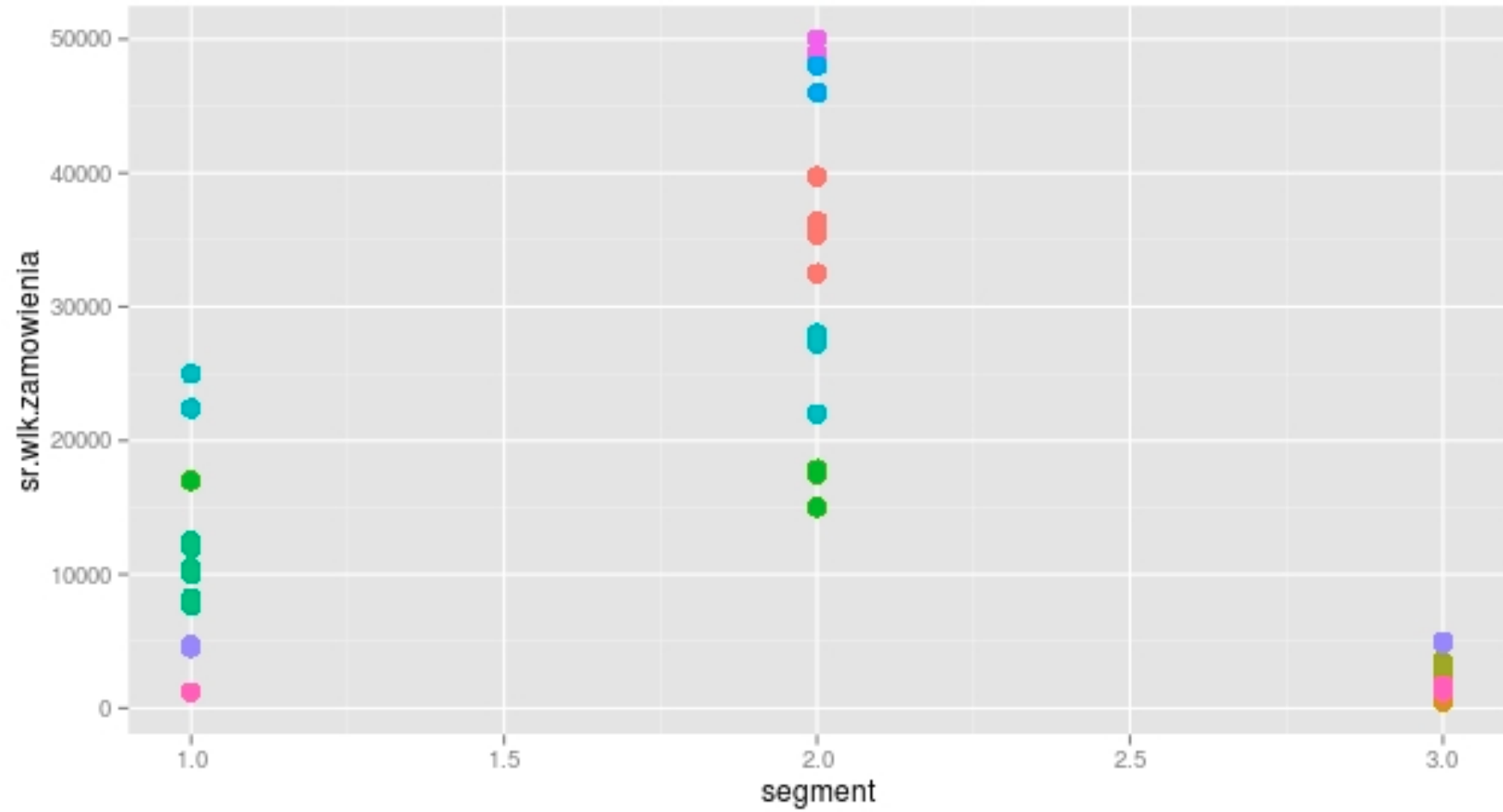




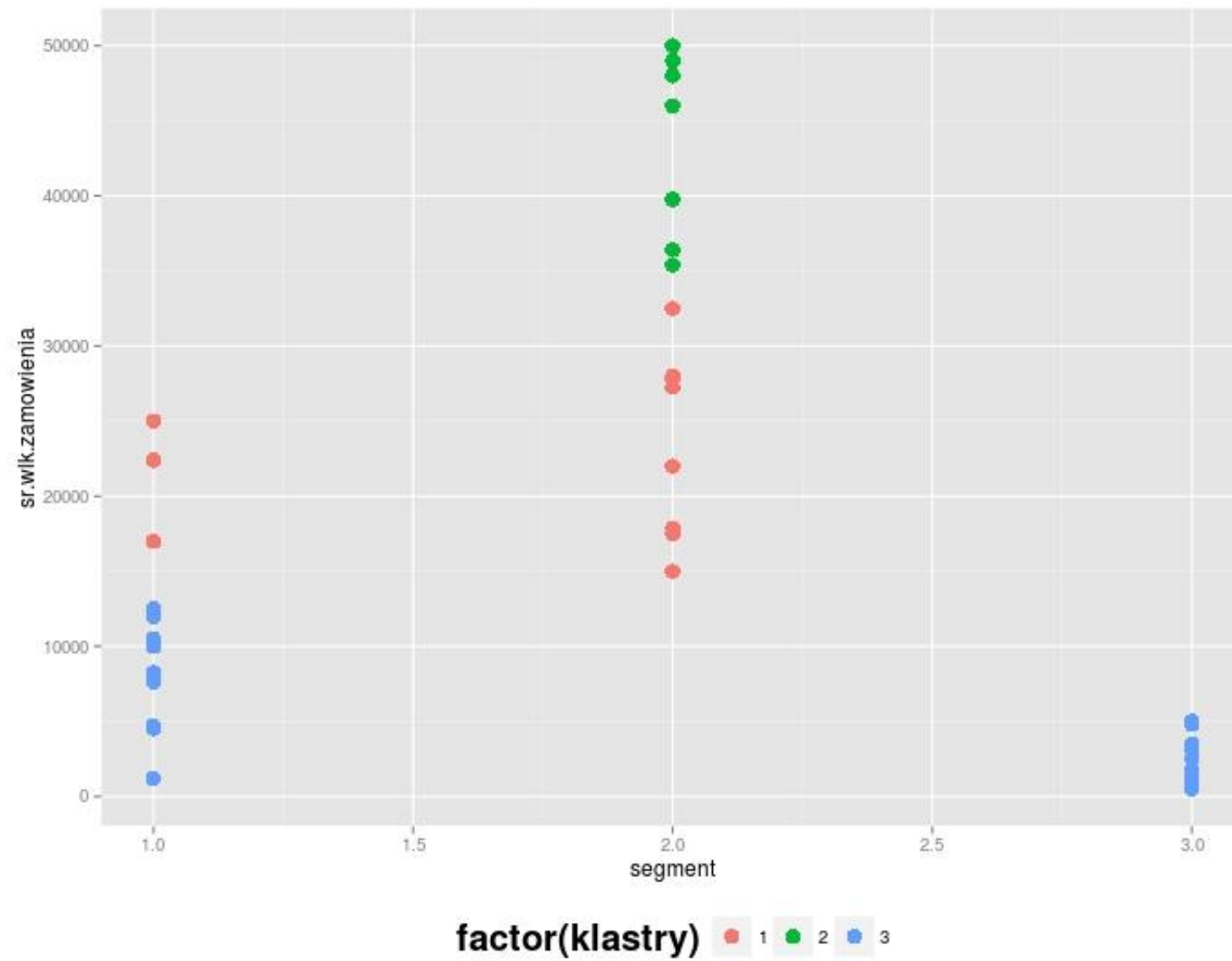


Sum of WithinSS Over Number of Clusters





factor(kmeans) 1 2 3 4 5 6 7 8 9 10



grupowanie umów
segmentacja klientów

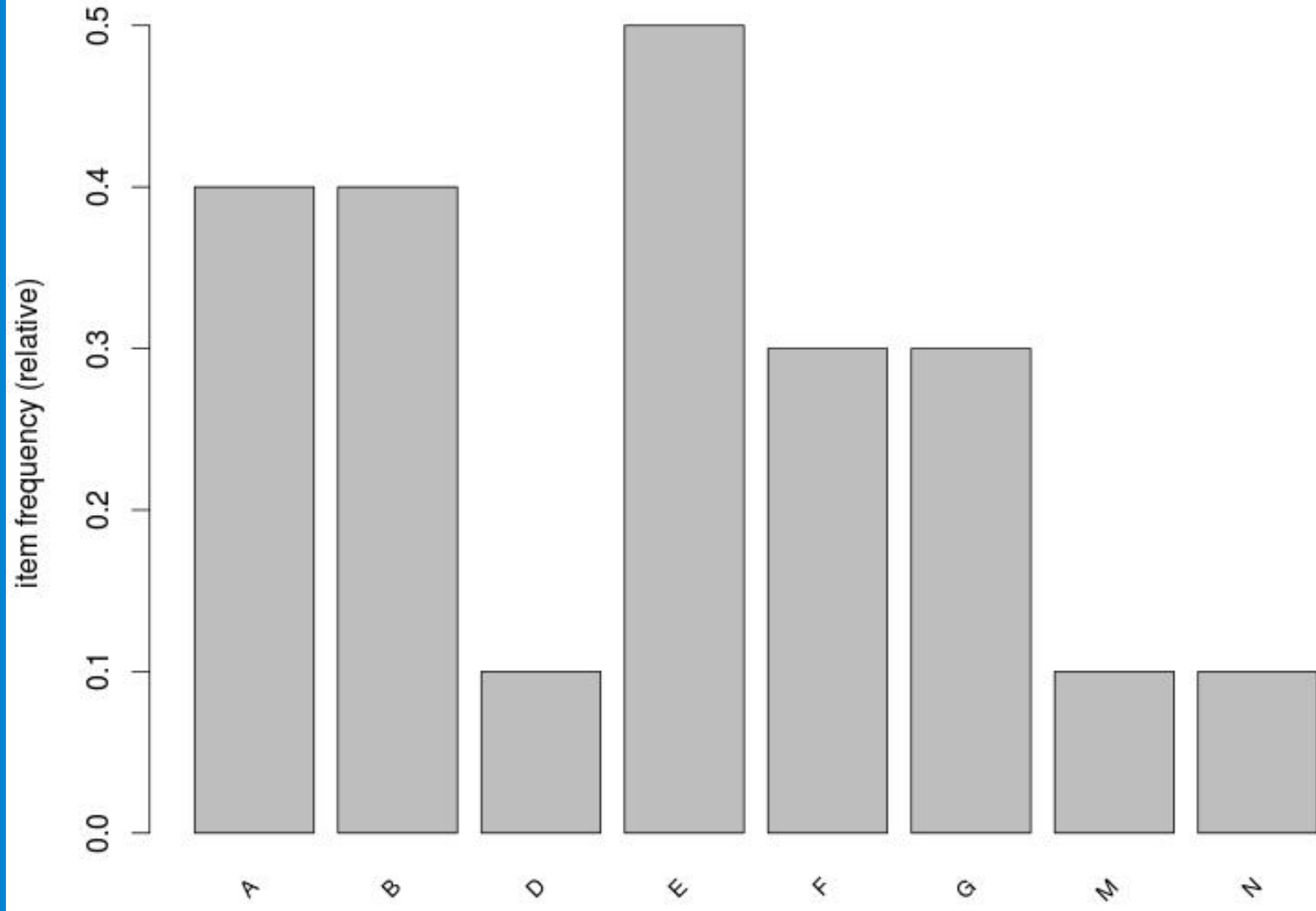
grupowanie umów
segmentacja klientów
definiowanie substytutów

grupowanie umów

segmentacja klientów

definiowanie substytutów

ocena stopnia konkurencji



Summary of the Measures of Interestingness:

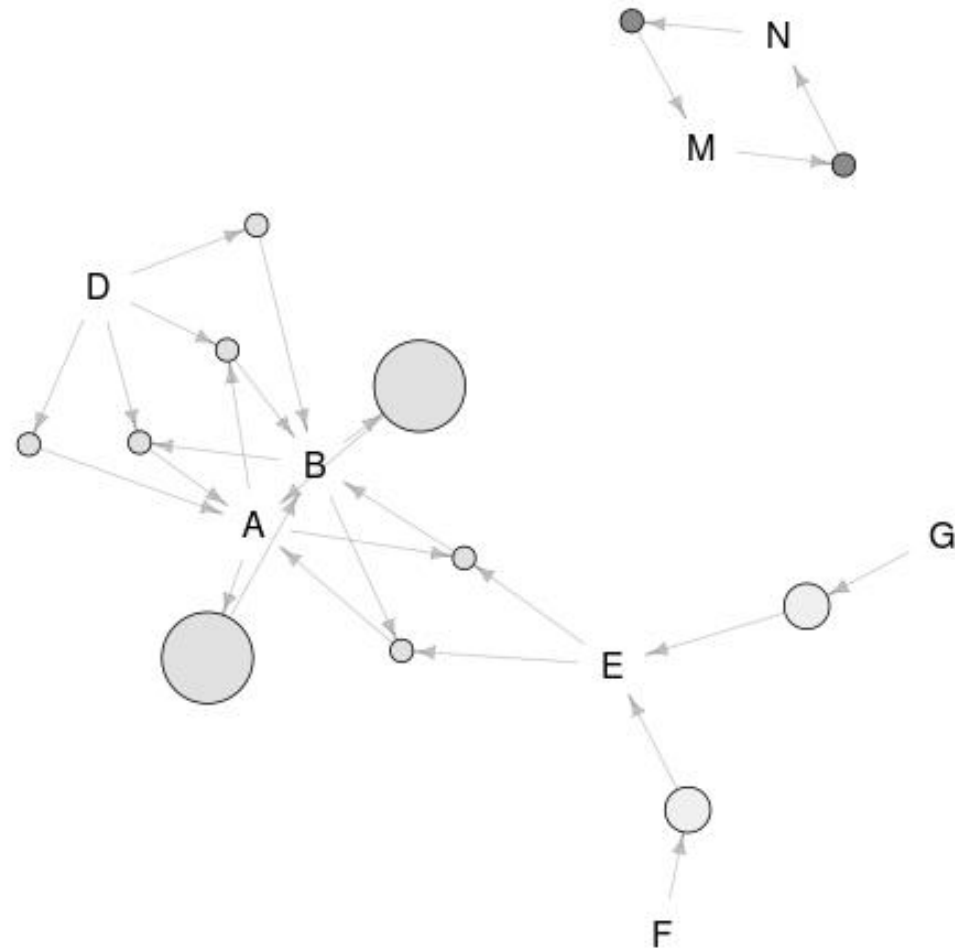
support	confidence	lift
Min. :0.1000	Min. :0.2000	Min. : 0.500
1st Qu.:0.1000	1st Qu.:0.2500	1st Qu.: 1.111
Median :0.1000	Median :0.5333	Median : 2.500
Mean :0.1417	Mean :0.6125	Mean : 2.398
3rd Qu.:0.1250	3rd Qu.:1.0000	3rd Qu.: 2.500
Max. :0.4000	Max. :1.0000	Max.: 10.000

All Rules

	lhs		rhs	support	confidence	lift
7	{B}	=>	{A}	0.4	1.0000000	2.500000
8	{A}	=>	{B}	0.4	1.0000000	2.500000
5	{F}	=>	{E}	0.2	0.6666667	1.333333
6	{G}	=>	{E}	0.2	0.6666667	1.333333
1	{M}	=>	{N}	0.1	1.0000000	10.000000
2	{N}	=>	{M}	0.1	1.0000000	10.000000
3	{D}	=>	{B}	0.1	1.0000000	2.500000
4	{D}	=>	{A}	0.1	1.0000000	2.500000
9	{B,D}	=>	{A}	0.1	1.0000000	2.500000
10	{A,D}	=>	{B}	0.1	1.0000000	2.500000
11	{B,E}	=>	{A}	0.1	1.0000000	2.500000
12	{A,E}	=>	{B}	0.1	1.0000000	2.500000

Graph for 12 rules

size: support (0.1 - 0.4)
color: lift (1.333 - 10)



=====

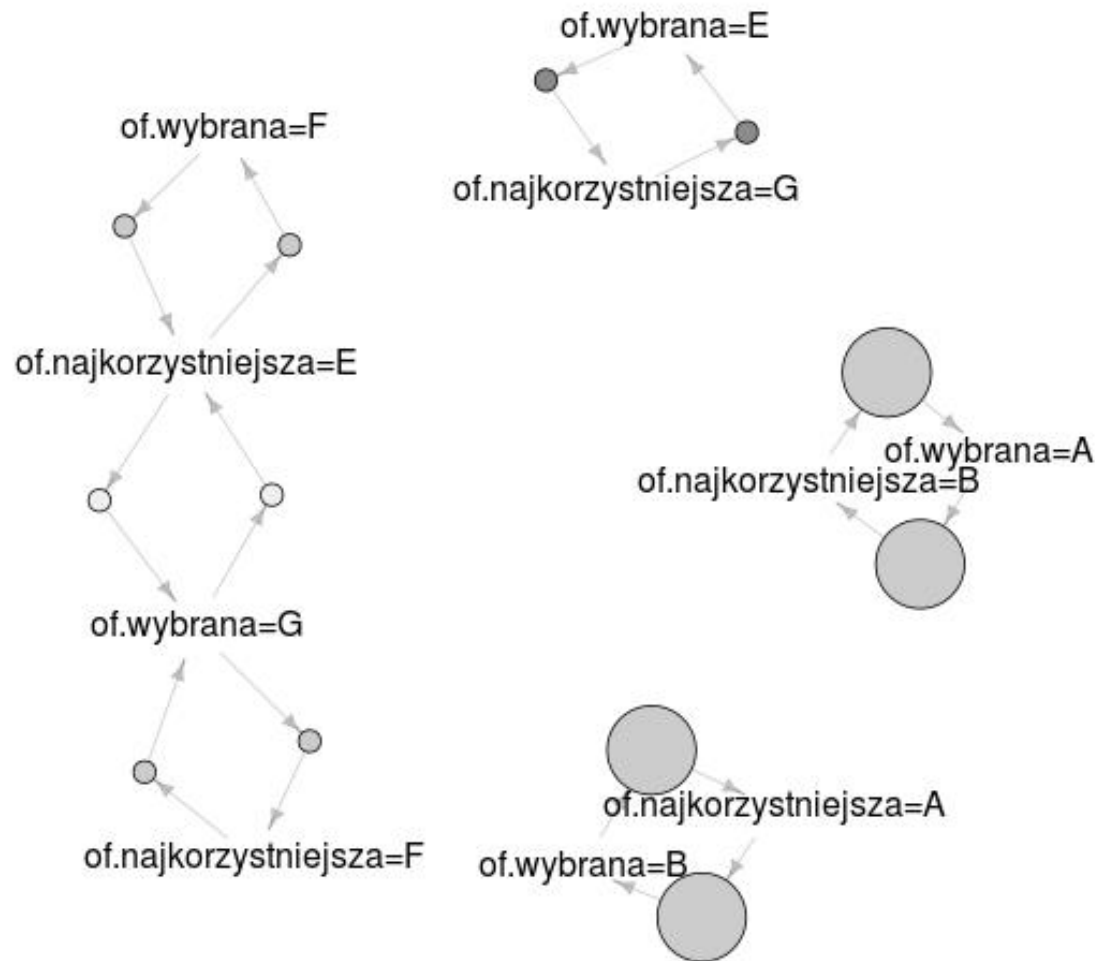
All Rules

	lhs	rhs	support	confidence	lift
7	{of.wybrana=B}	=> {of.najkorzystniejsza=A}	0.250	1.0	4
8	{of.najkorzystniejsza=A}	=> {of.wybrana=B}	0.250	1.0	4
9	{of.najkorzystniejsza=B}	=> {of.wybrana=A}	0.250	1.0	4
10	{of.wybrana=A}	=> {of.najkorzystniejsza=B}	0.250	1.0	4
1	{of.najkorzystniejsza=F}	=> {of.wybrana=G}	0.125	1.0	4
2	{of.wybrana=G}	=> {of.najkorzystniejsza=F}	0.125	0.5	4
3	{of.najkorzystniejsza=G}	=> {of.wybrana=E}	0.125	1.0	8
4	{of.wybrana=E}	=> {of.najkorzystniejsza=G}	0.125	1.0	8
5	{of.wybrana=F}	=> {of.najkorzystniejsza=E}	0.125	1.0	4
6	{of.najkorzystniejsza=E}	=> {of.wybrana=F}	0.125	0.5	4
11	{of.wybrana=G}	=> {of.najkorzystniejsza=E}	0.125	0.5	2
12	{of.najkorzystniejsza=E}	=> {of.wybrana=G}	0.125	0.5	2

Graph for 12 rules

size: support (0.125 - 0.25)

color: lift (2 - 8)



rozstawianie i wycofywanie ofert

rozstawianie i wycofywanie ofert

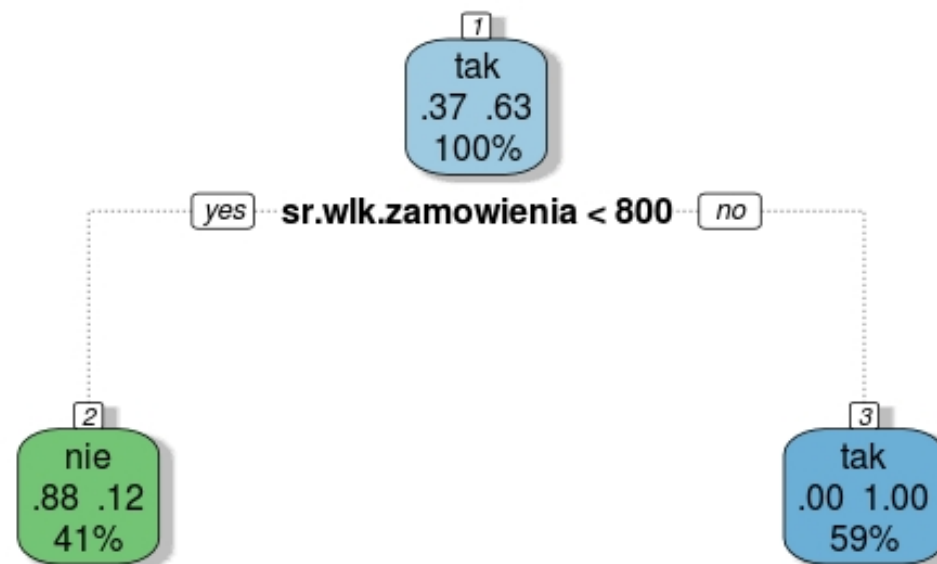
relacje między przedsiębiorcami

rozstawianie i wycofywanie ofert

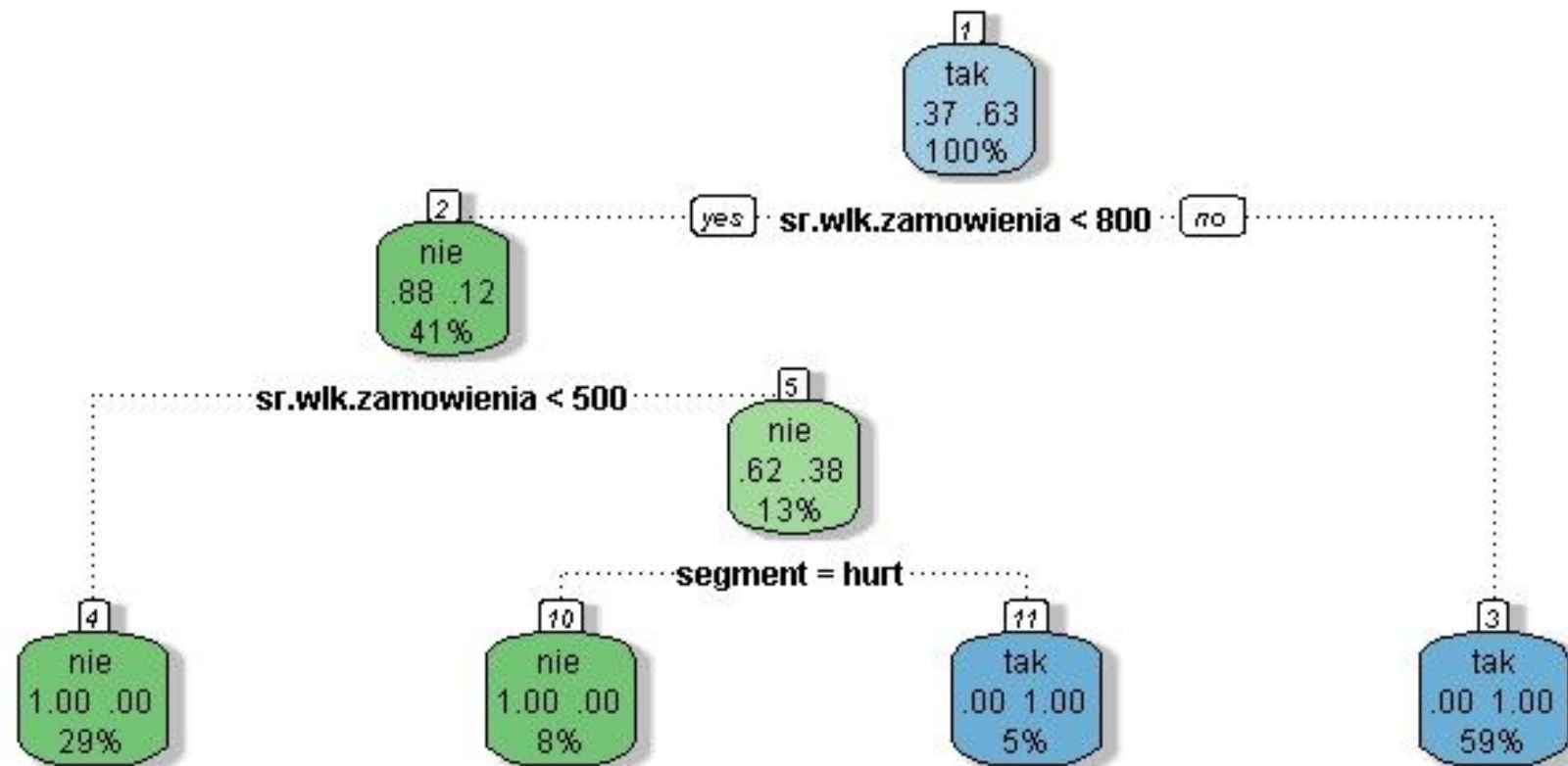
relacje między przedsiębiorcami

ocena stopnia substytucji towarów

rozstawianie i wycofywanie ofert
relacje między przedsiębiorcami
ocena stopnia substytucji towarów
budowa koszyków towarów



Rattle 2015-paź-11 19:46:54 ubuntu



Rattle 2015-wrz-27 18:10:33 user

weryfikacja polityki handlowej

weryfikacja polityki handlowej
analiza praktyk wykluczających

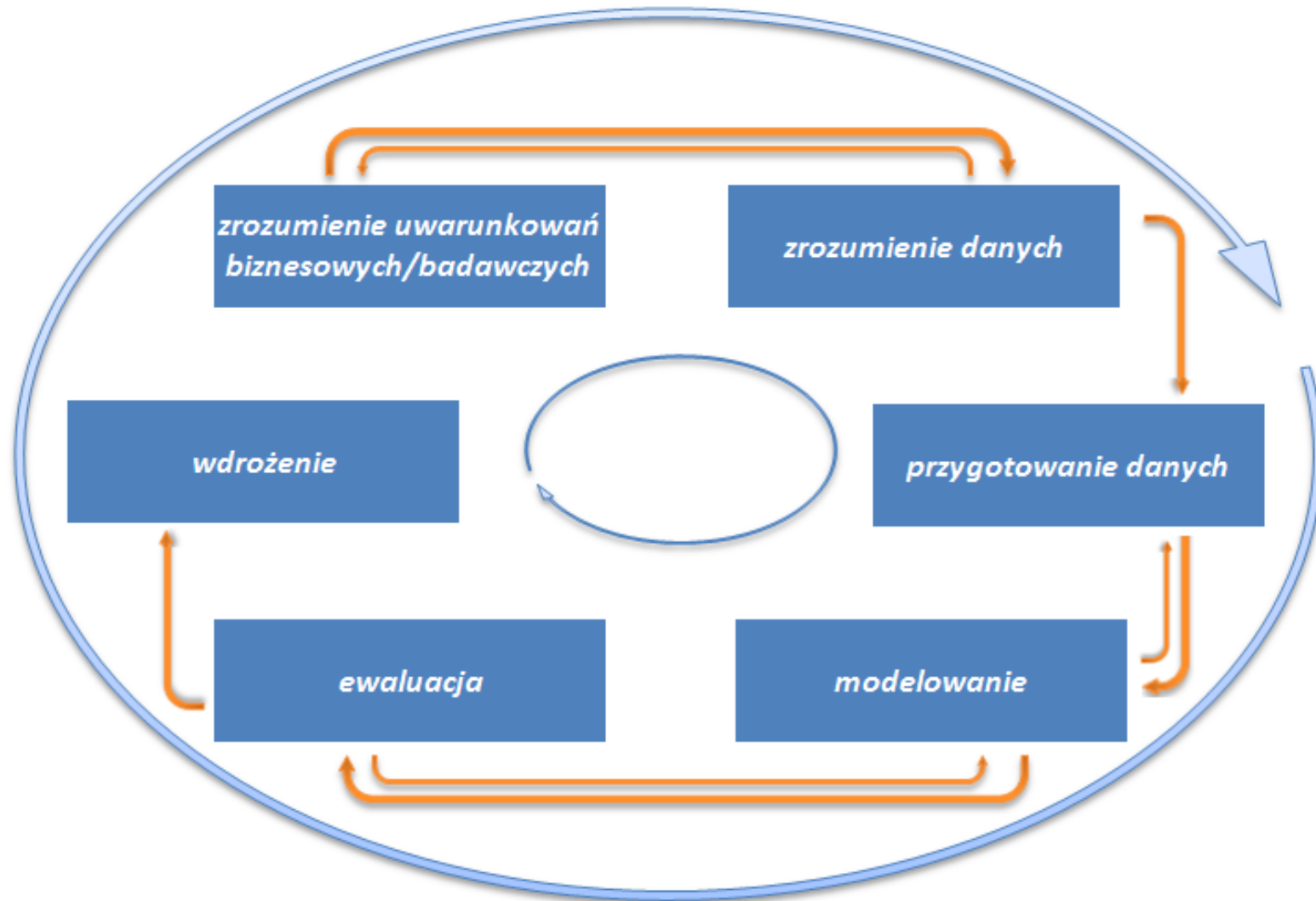
weryfikacja polityki handlowej
analiza praktyk wykluczających
funkcjonowanie sieci dystrybucyjnych

weryfikacja polityki handlowej

analiza praktyki decyzyjnej

funkcjonowanie sieci dystrybucyjnych

analiza praktyk wykluczających



grafiki

slajd 4

<http://ftalphaville.ft.com/2013/01/14/1333552/chinas-non-conforming-gdp-growth/>

slajd 5

https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Clustering/Hybrid_Hierarchical_Clustering

slajd 6

<https://alliance.seas.upenn.edu/~cis520/wiki/index.php?n=Lectures.DecisionTrees>

slajd 7

http://www.allanalytics.com/author.asp?section_id=2037

slajdy 9 – 11

Abrantez-Metz A. R., S.B. Villas-Boas, G. G. Judge, Tracking the LIBOR Rate, 03.11.2013, CUDARE Working Papers,

slajd 31

https://pl.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

font

LATO - Łukasz Dziejczak

<http://www.latofonts.com/>