How to display data badly

Karl W. Broman
Department of Biostatistics

http://www.biostat.jhsph.edu/~kbroman

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Using Microsoft Excel to obscure your data and annoy your readers

Karl W Broman
Department of Biostatistics

http://www.biostat.jhsph.edu/~kbroman

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Inspiration

This lecture was inspired by

H Wainer (1984) How to display data badly. American Statistician 38(2):137-147

Dr. Wainer was the first to elucidate the principles of the bad display of data.

The now widespread use of Microsoft Excel has resulted in remarkable advances in the field.

General principles

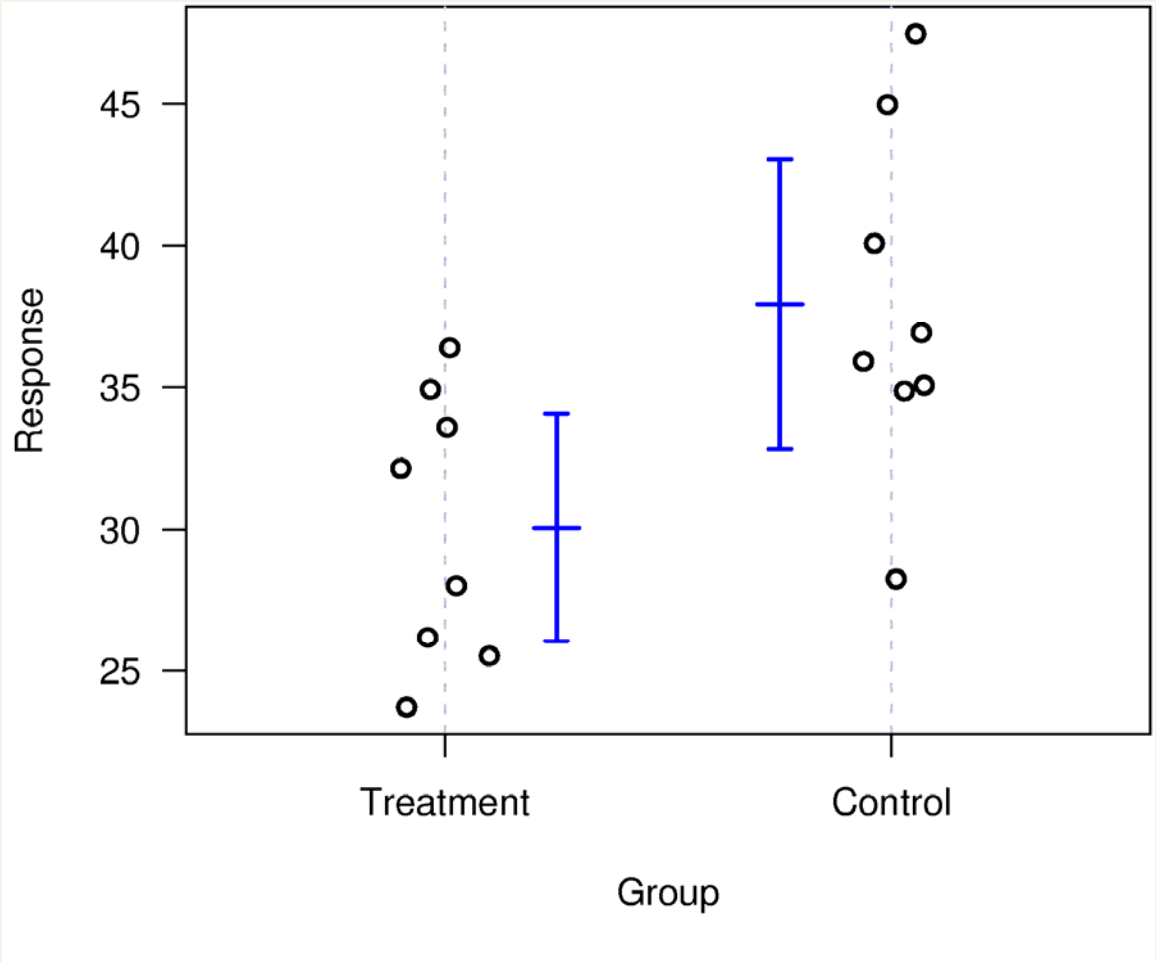
The aim of good data graphics:

Display data accurately and clearly.

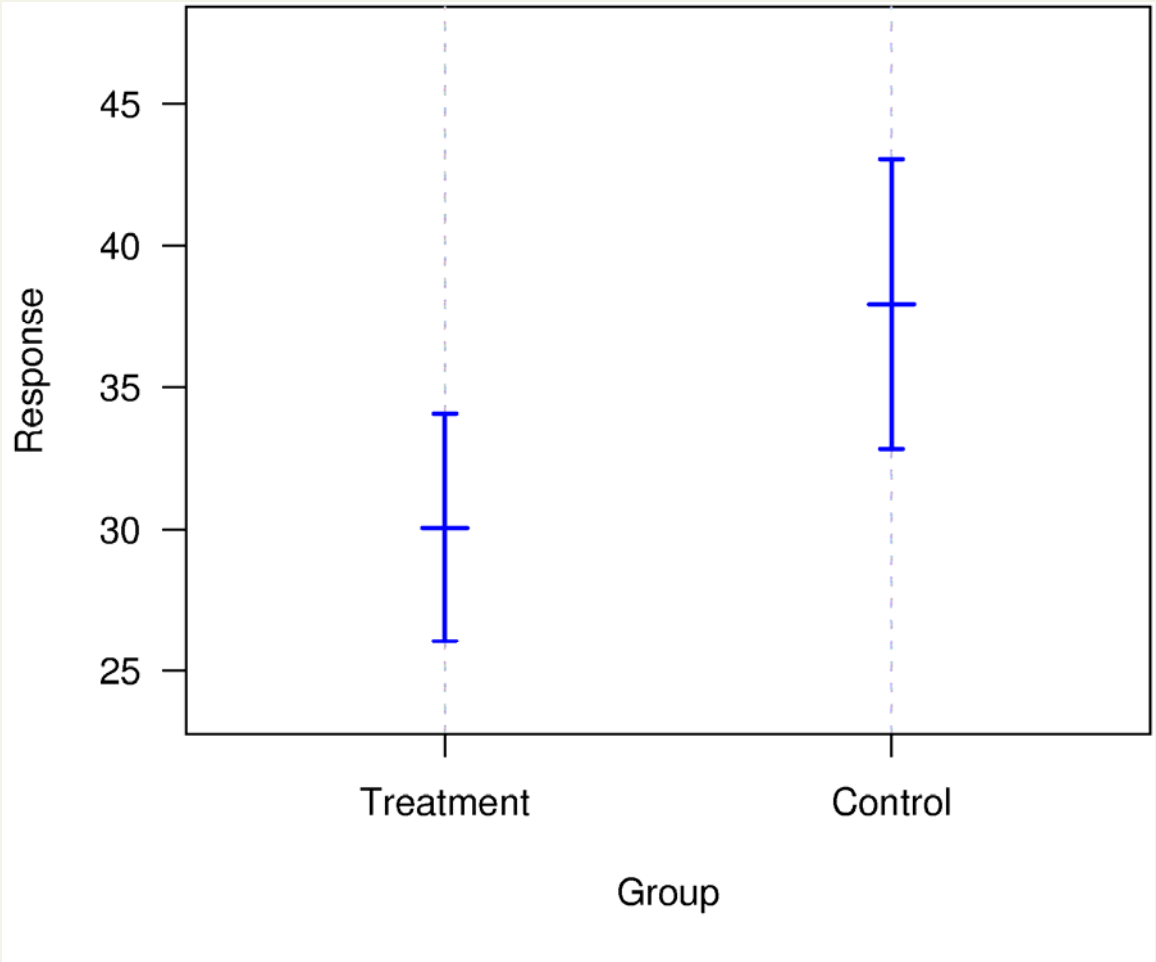
Some rules for displaying data badly:

- Display as little information as possible.
- Obscure what you do show (with chart junk).
- Use pseudo-3d and color gratuitously.
- Make a pie chart (preferably in color and 3d).
- Use a poorly chosen scale.
- Ignore sig figs.

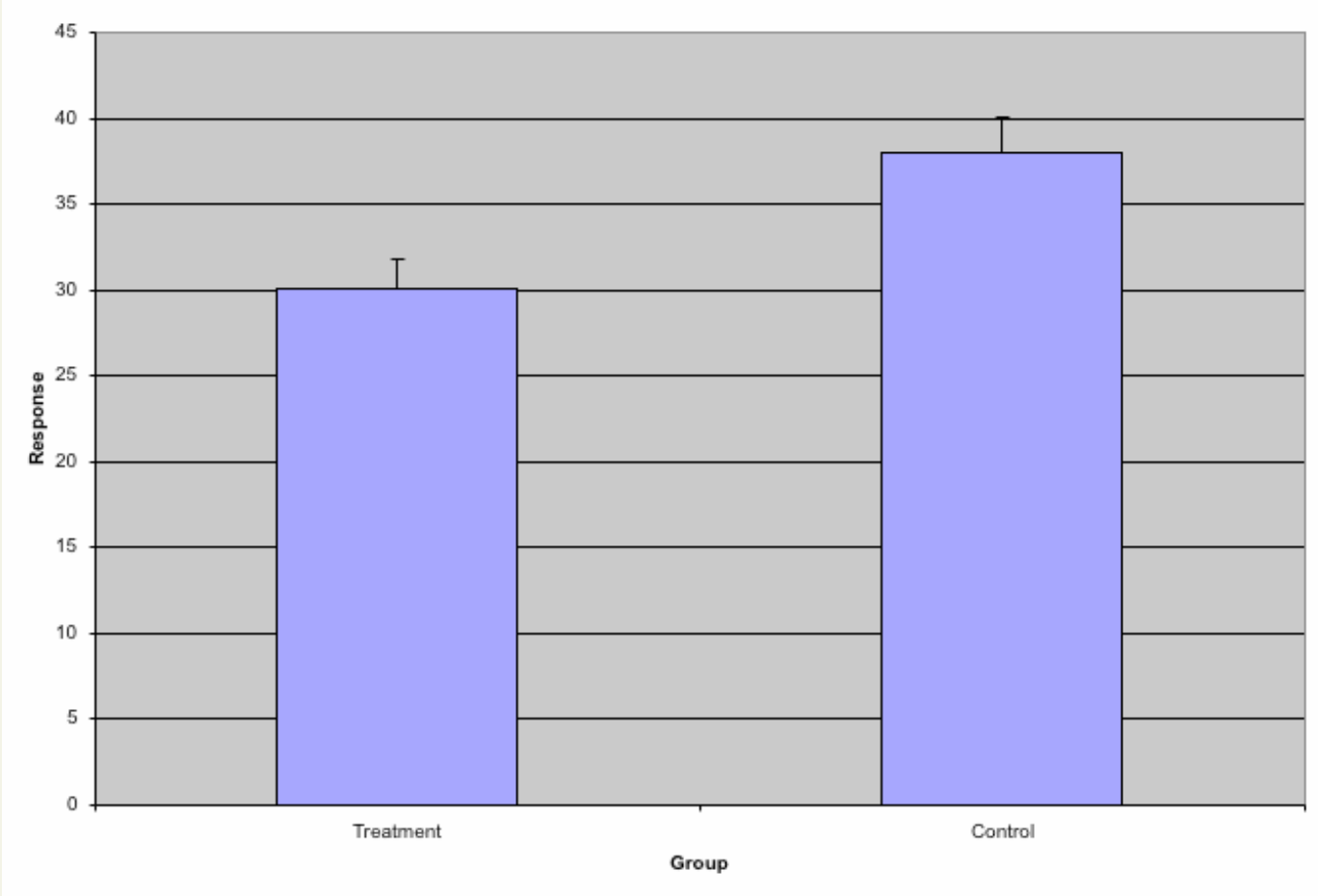
Example 1



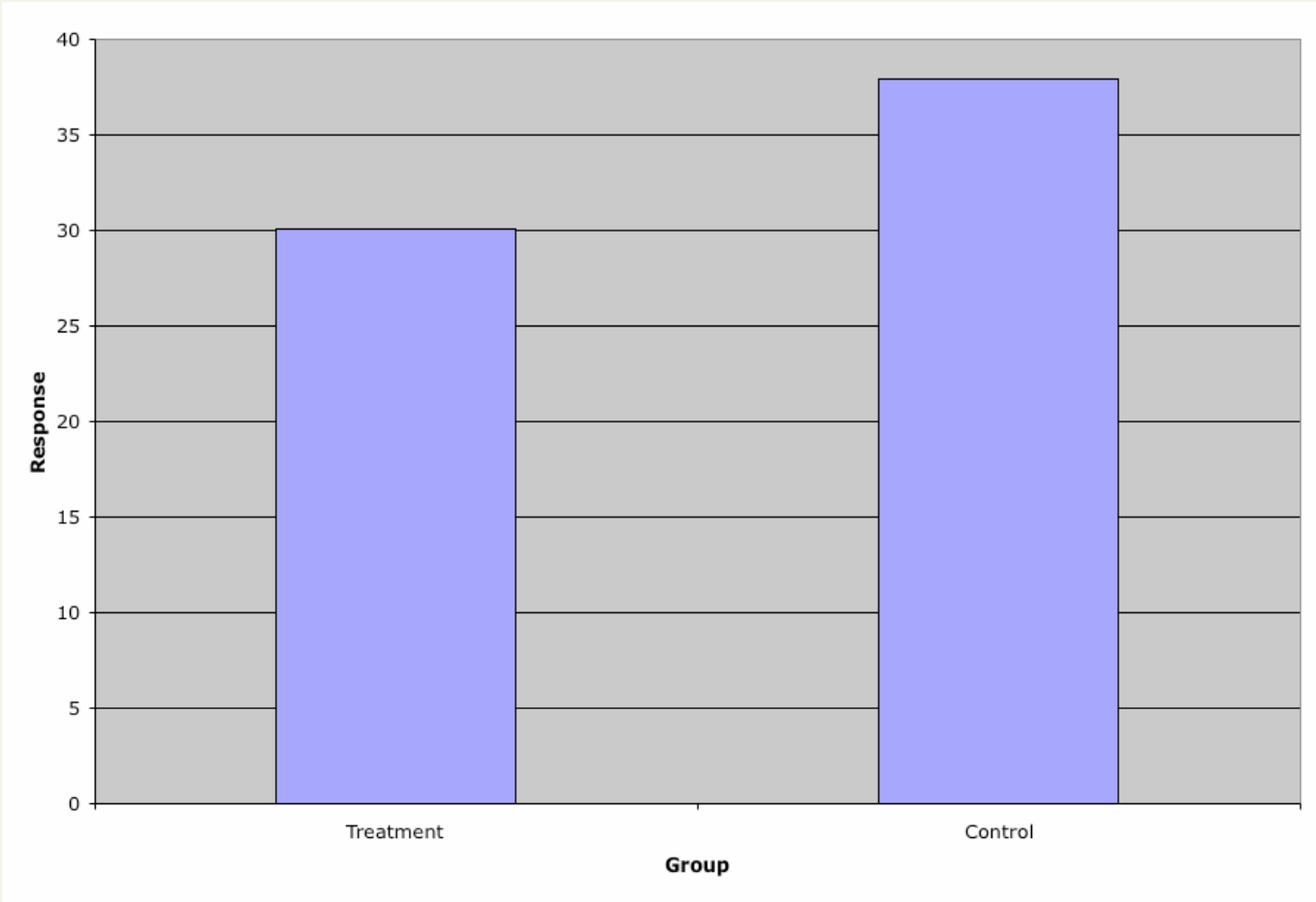
Example 1



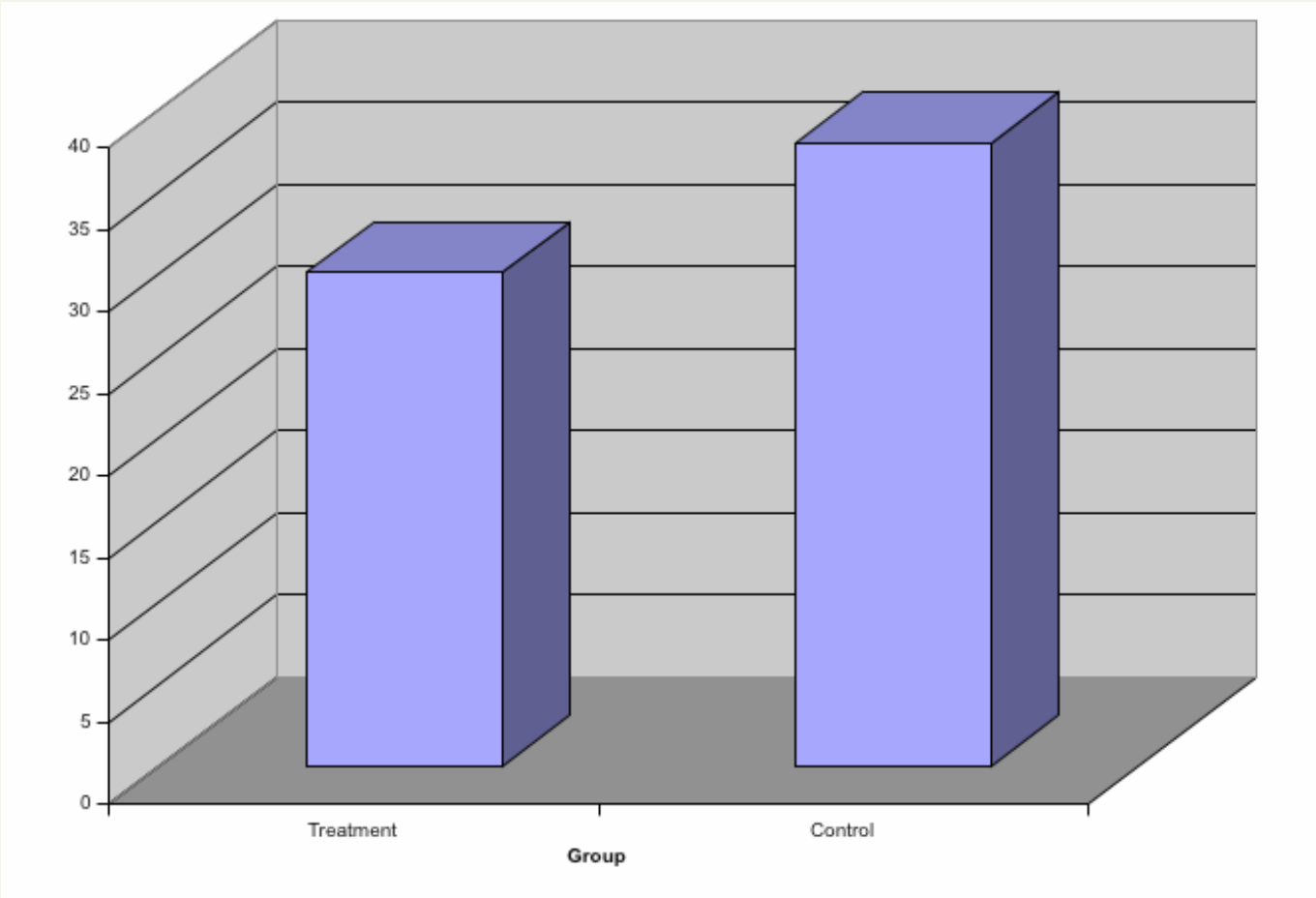
Example 1



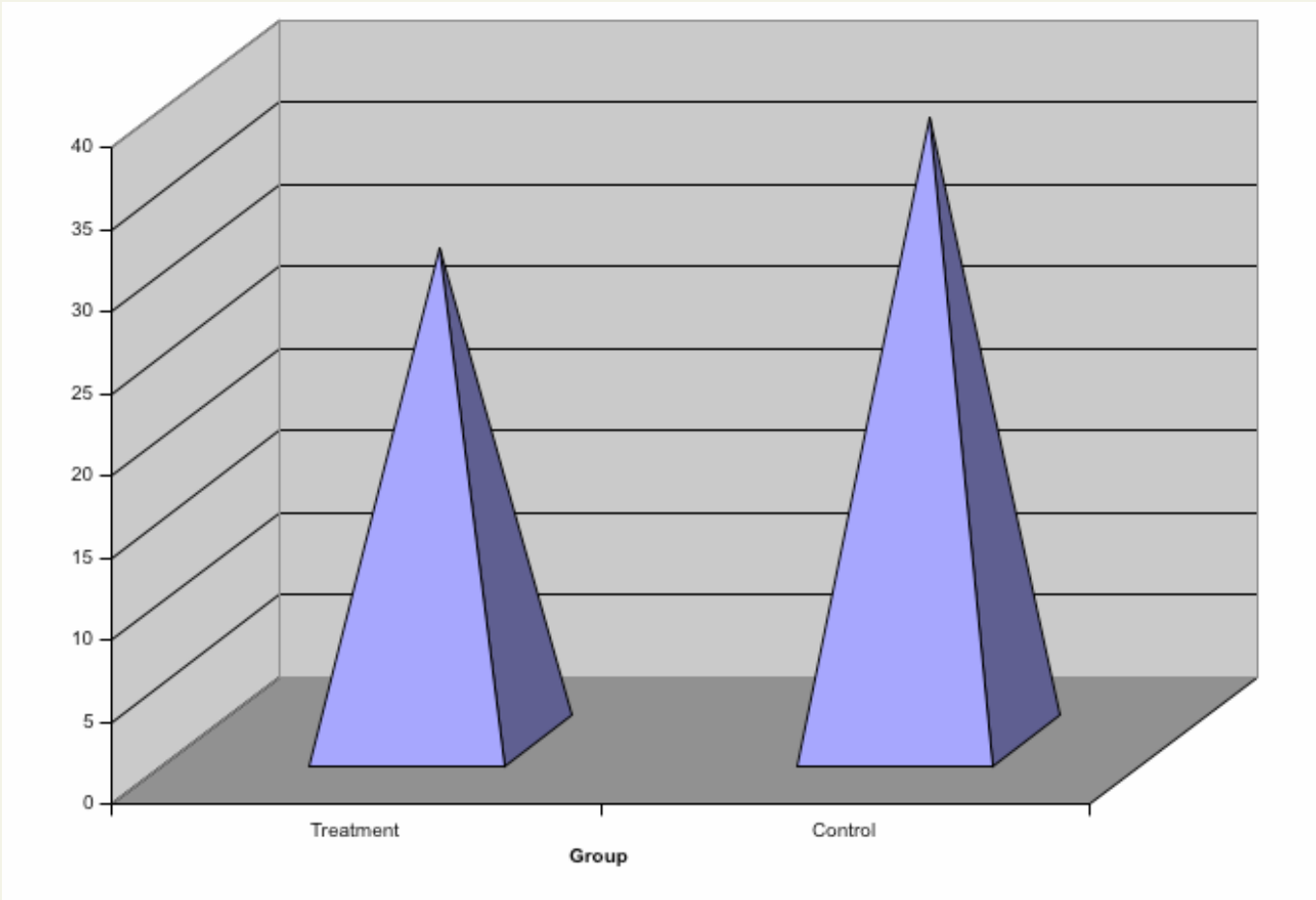
Example 1



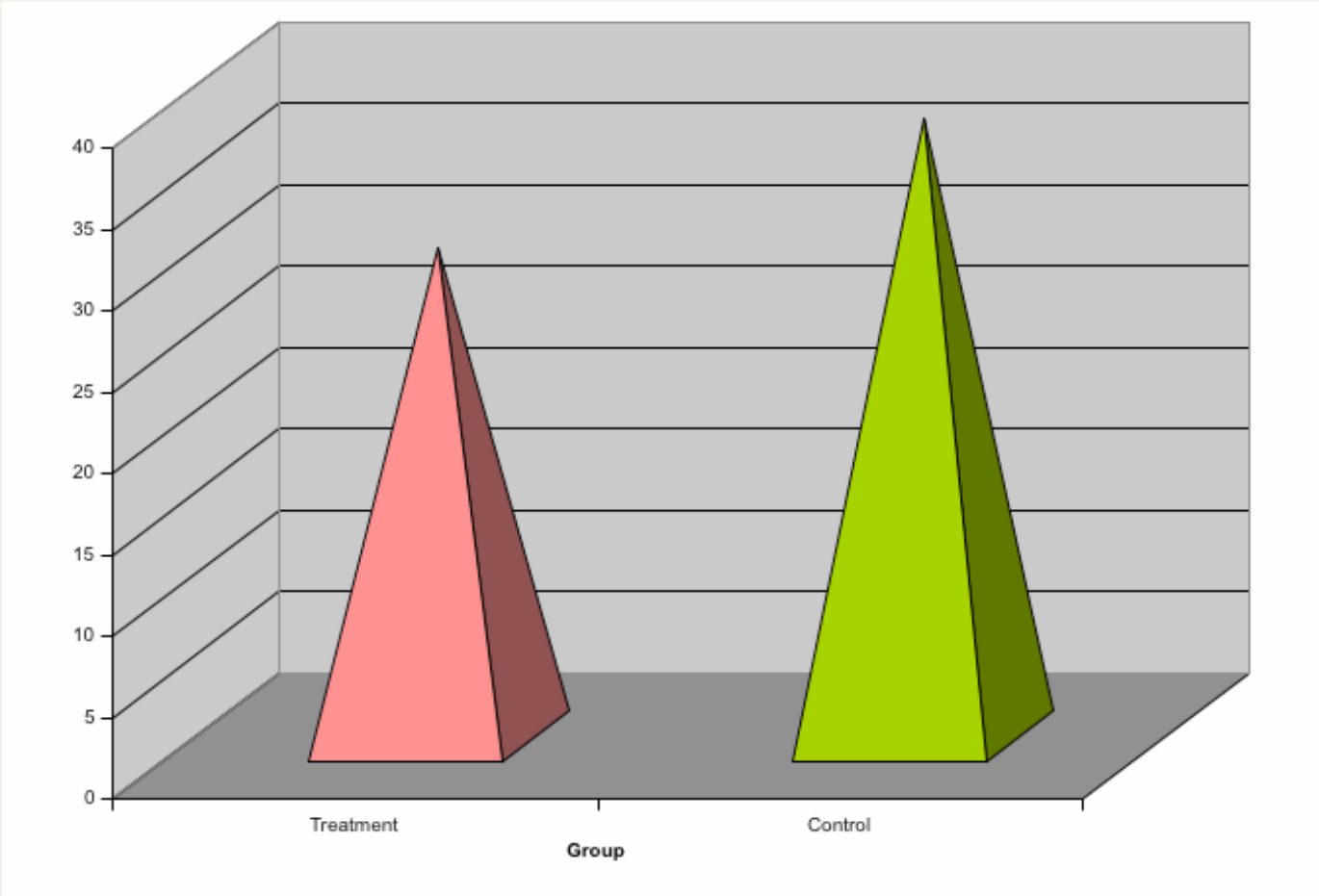
Example 1



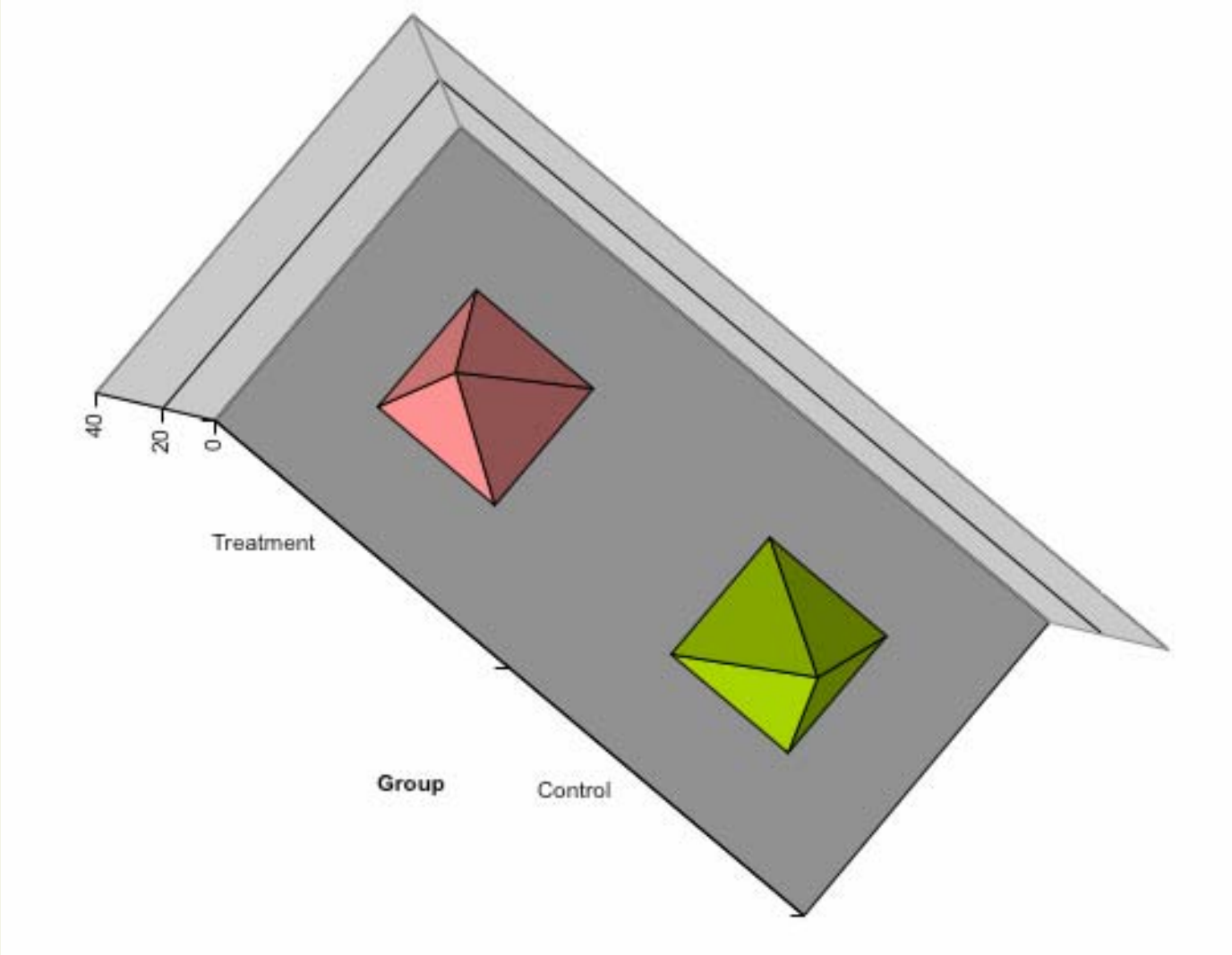
Example 1



Example 1



Example 1

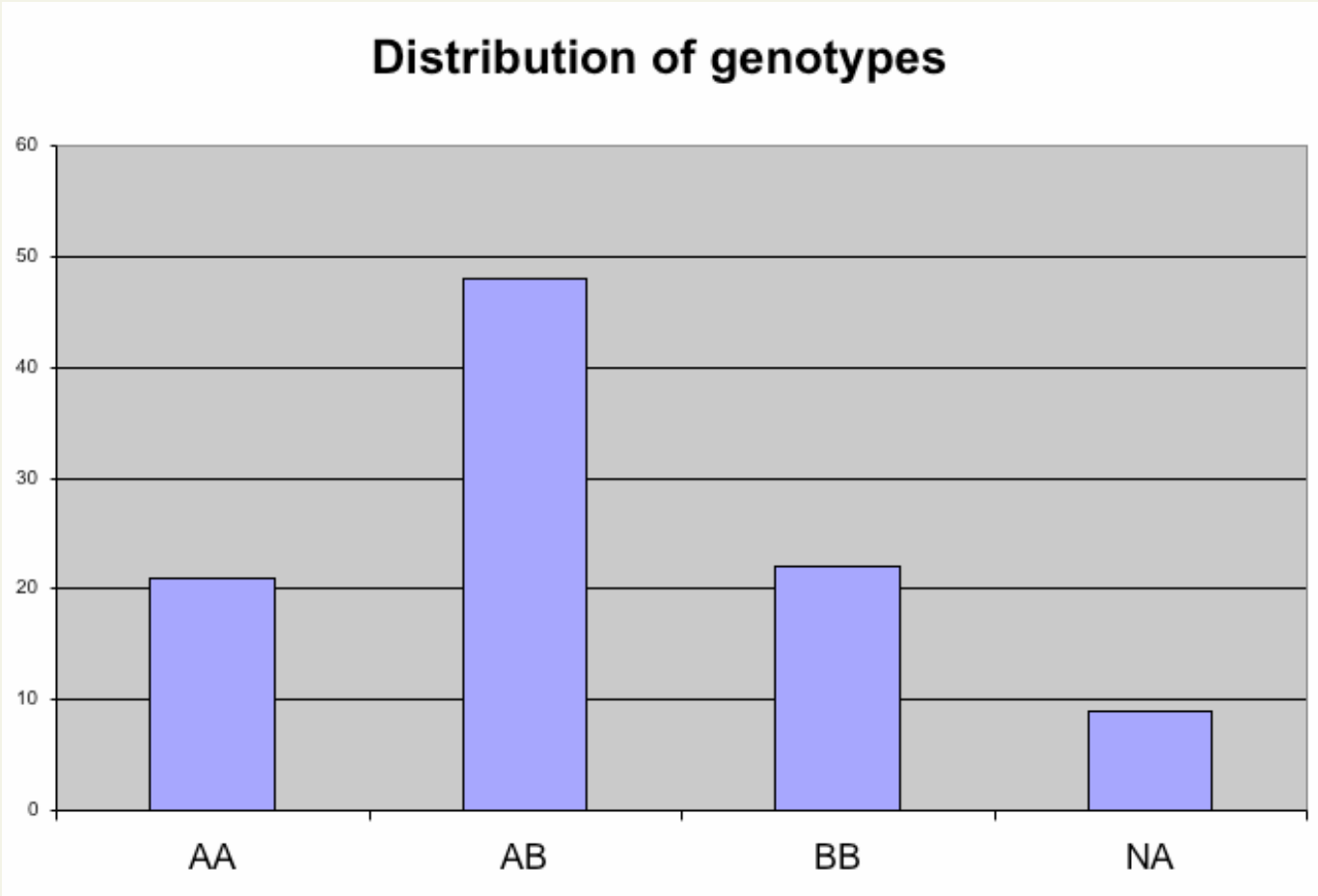


Example 2

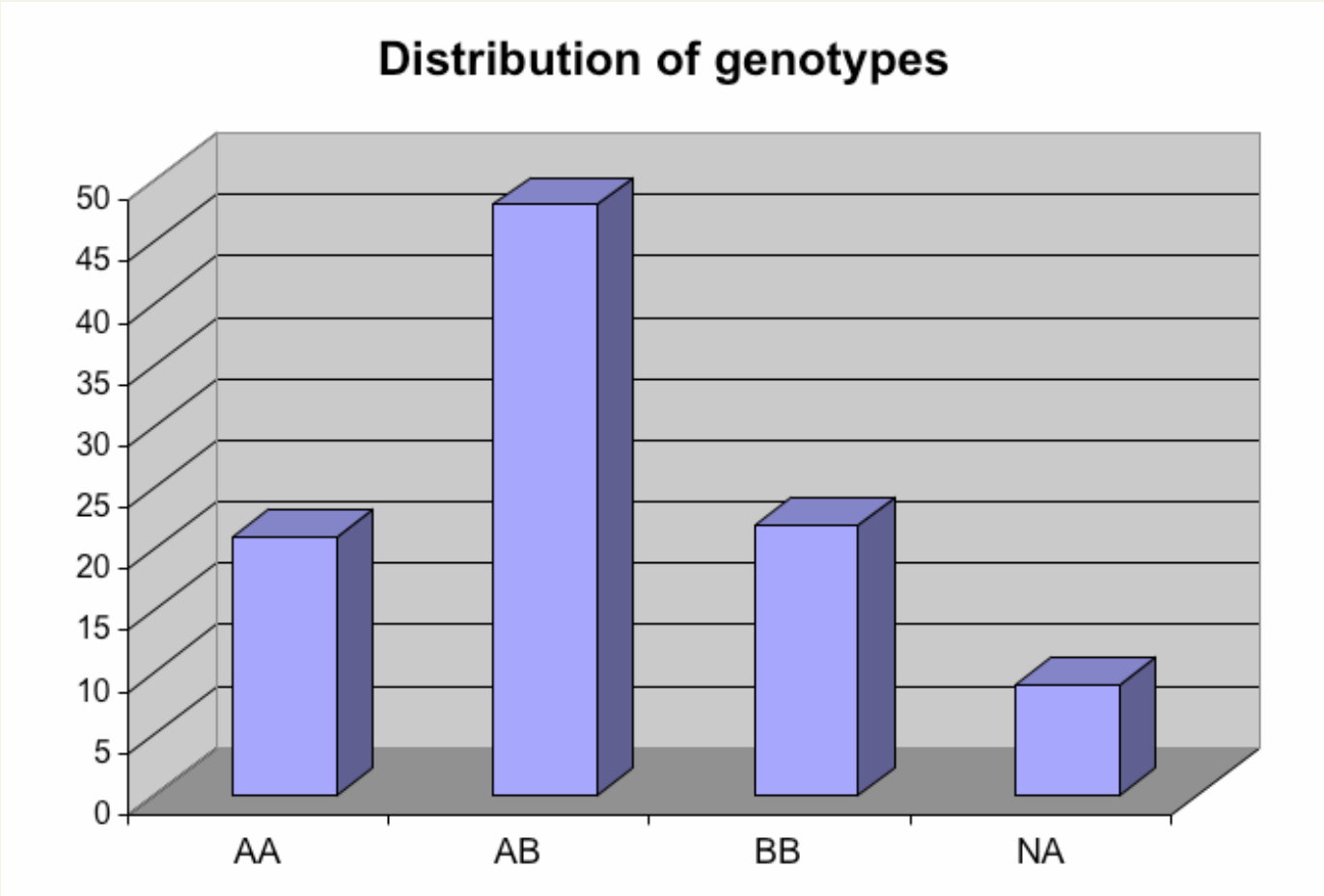
Distribution of genotypes

AA	21%
AB	48%
BB	22%
missing	9%

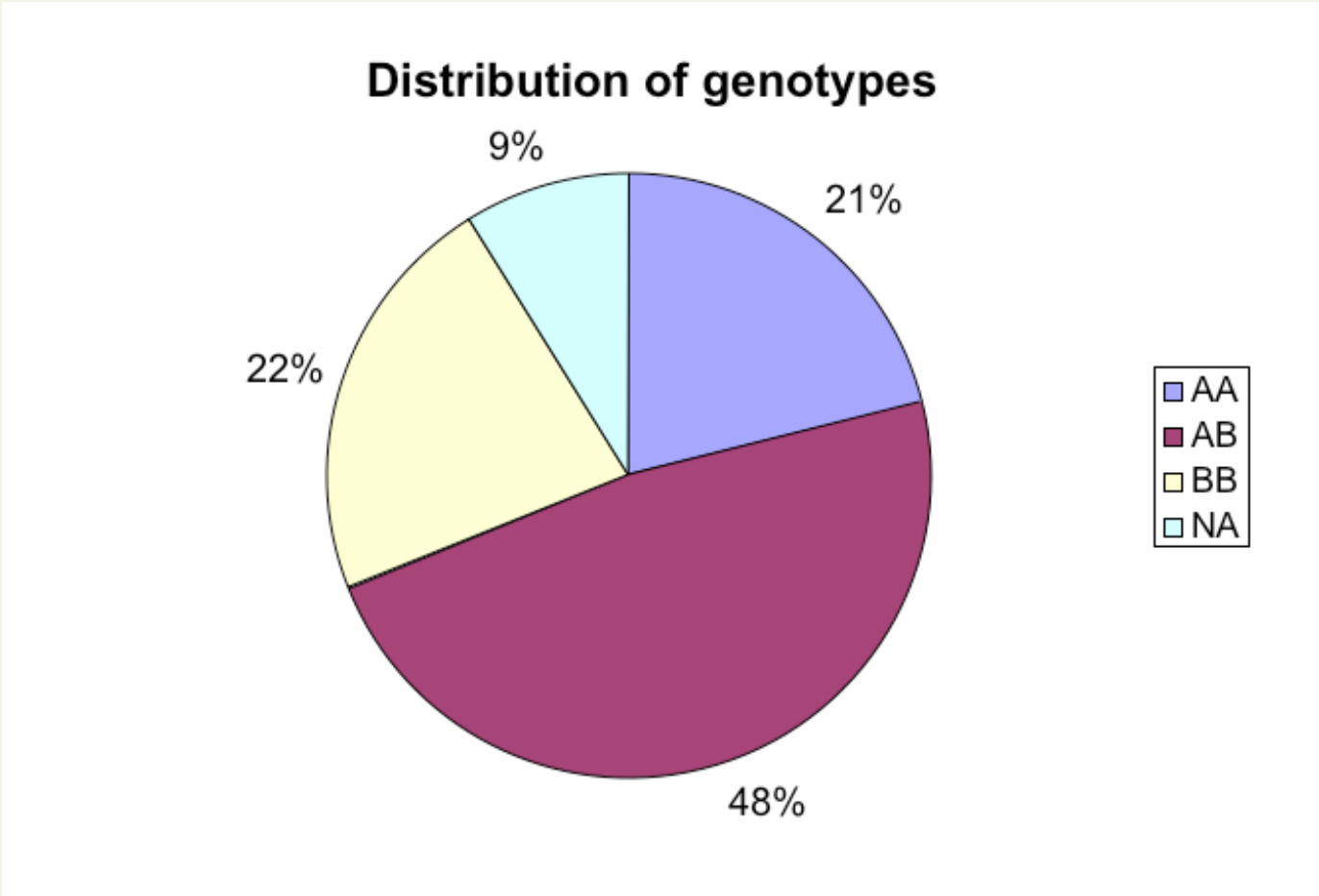
Example 2



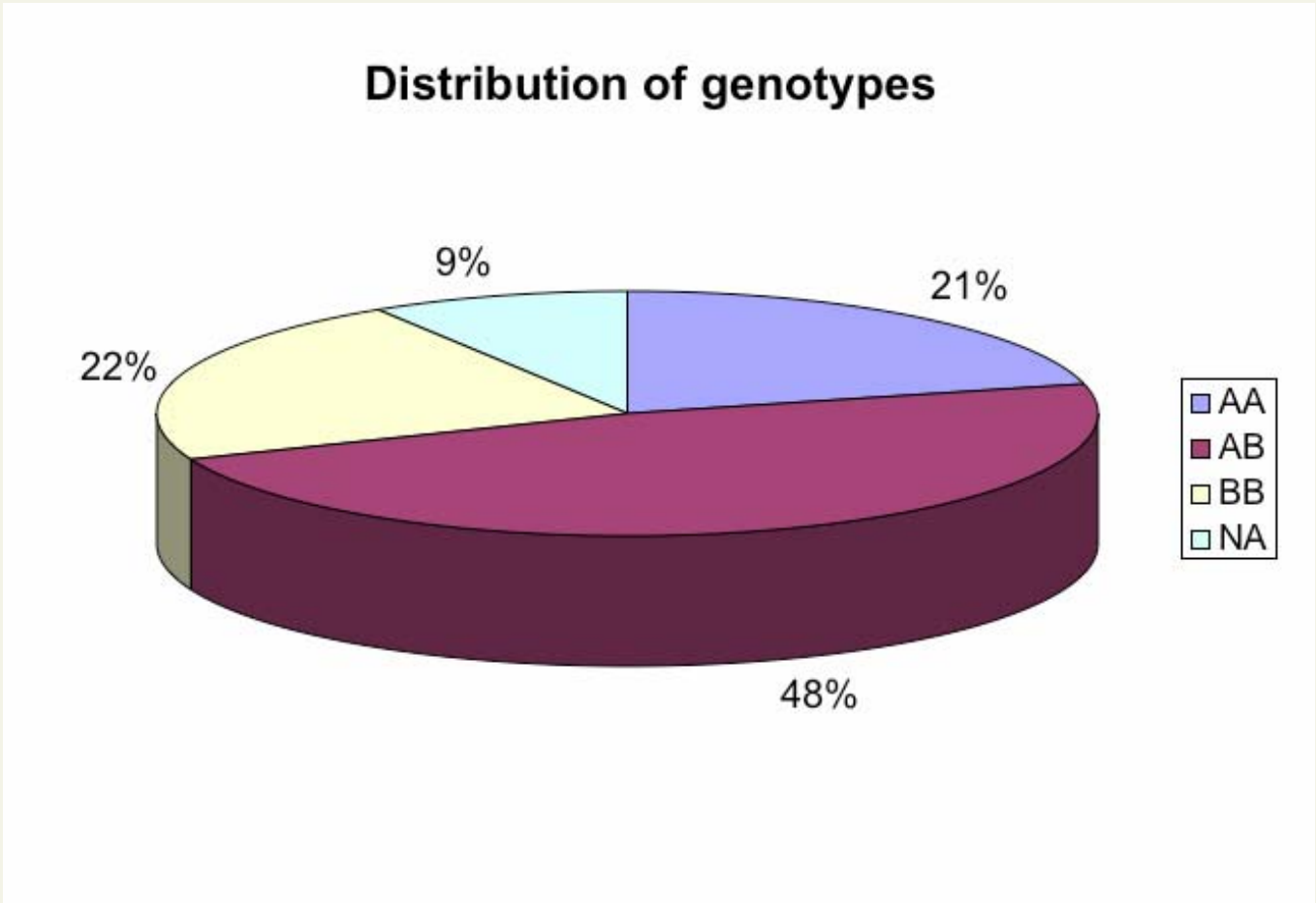
Example 2


Distribution of genotypes
50
45
40
35
30
25
20
15
10
5
0
AA
AB
BB
NA

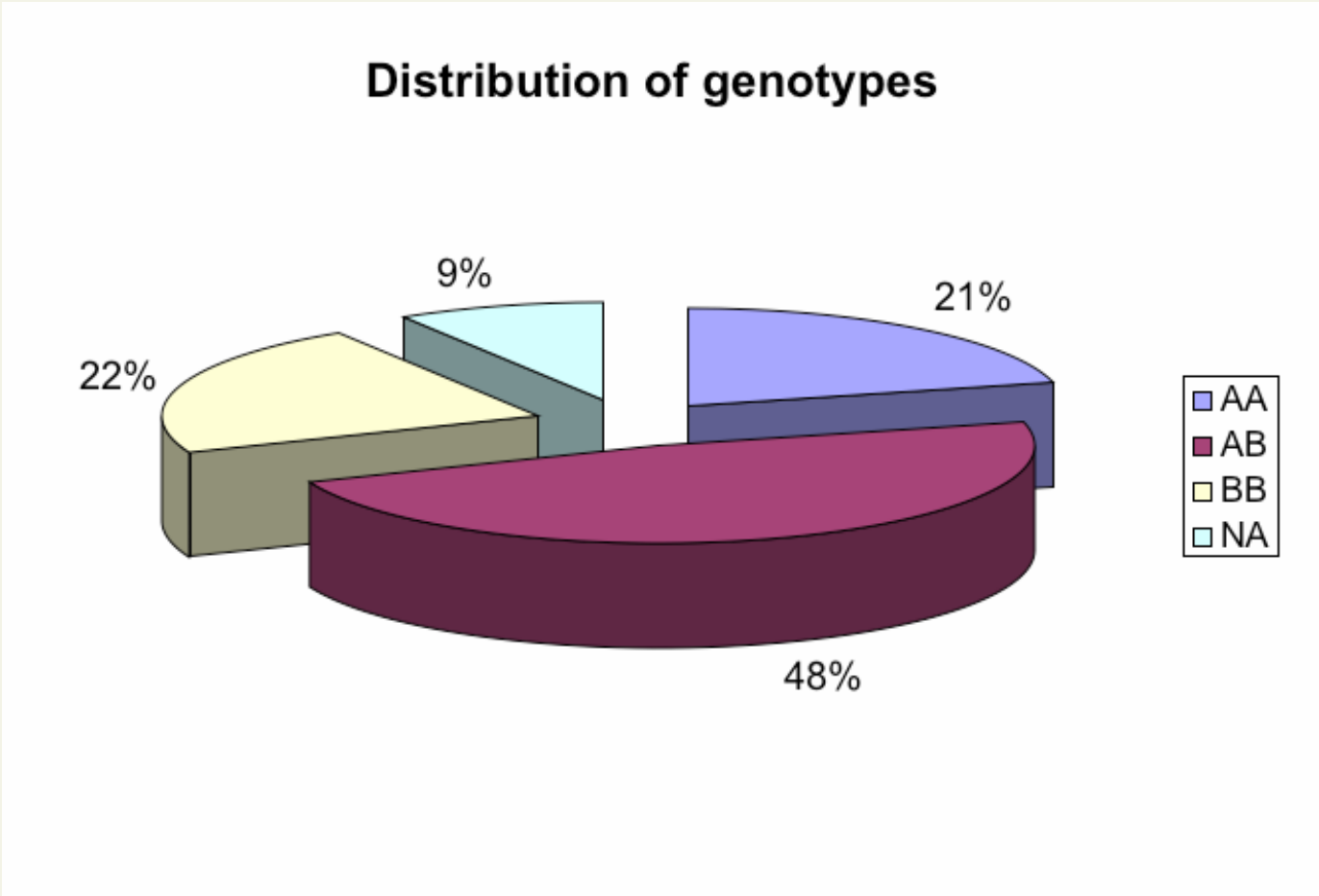
Example 2



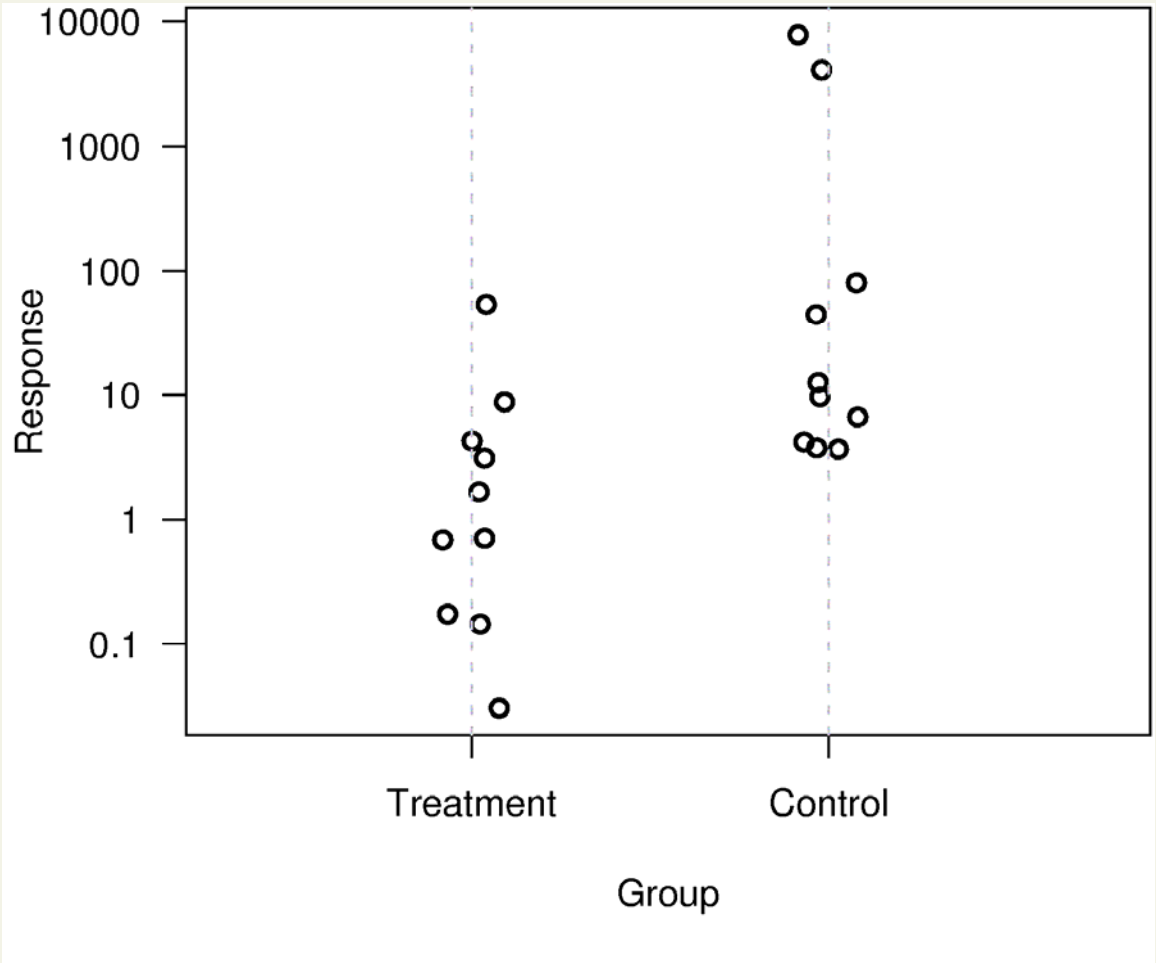
Example 2



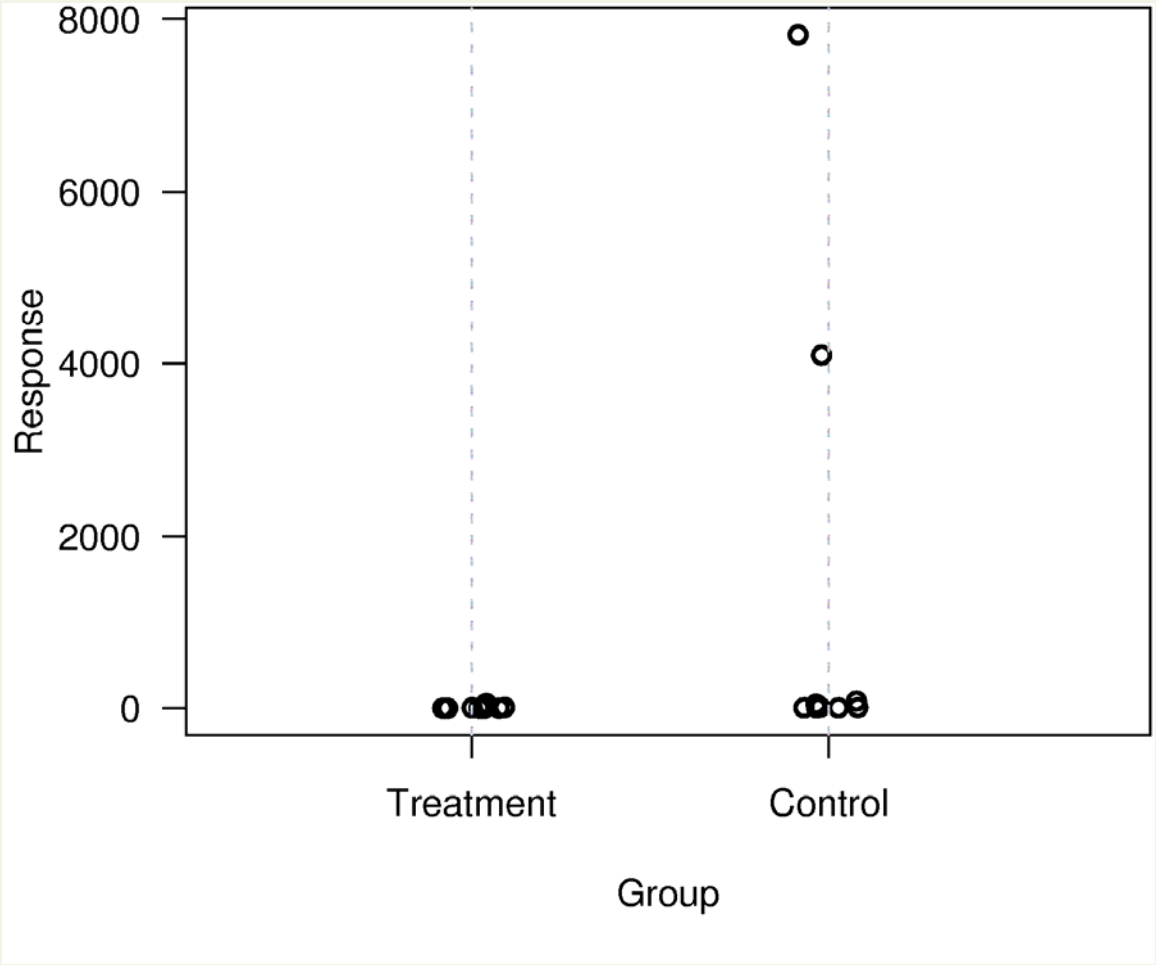
Example 2



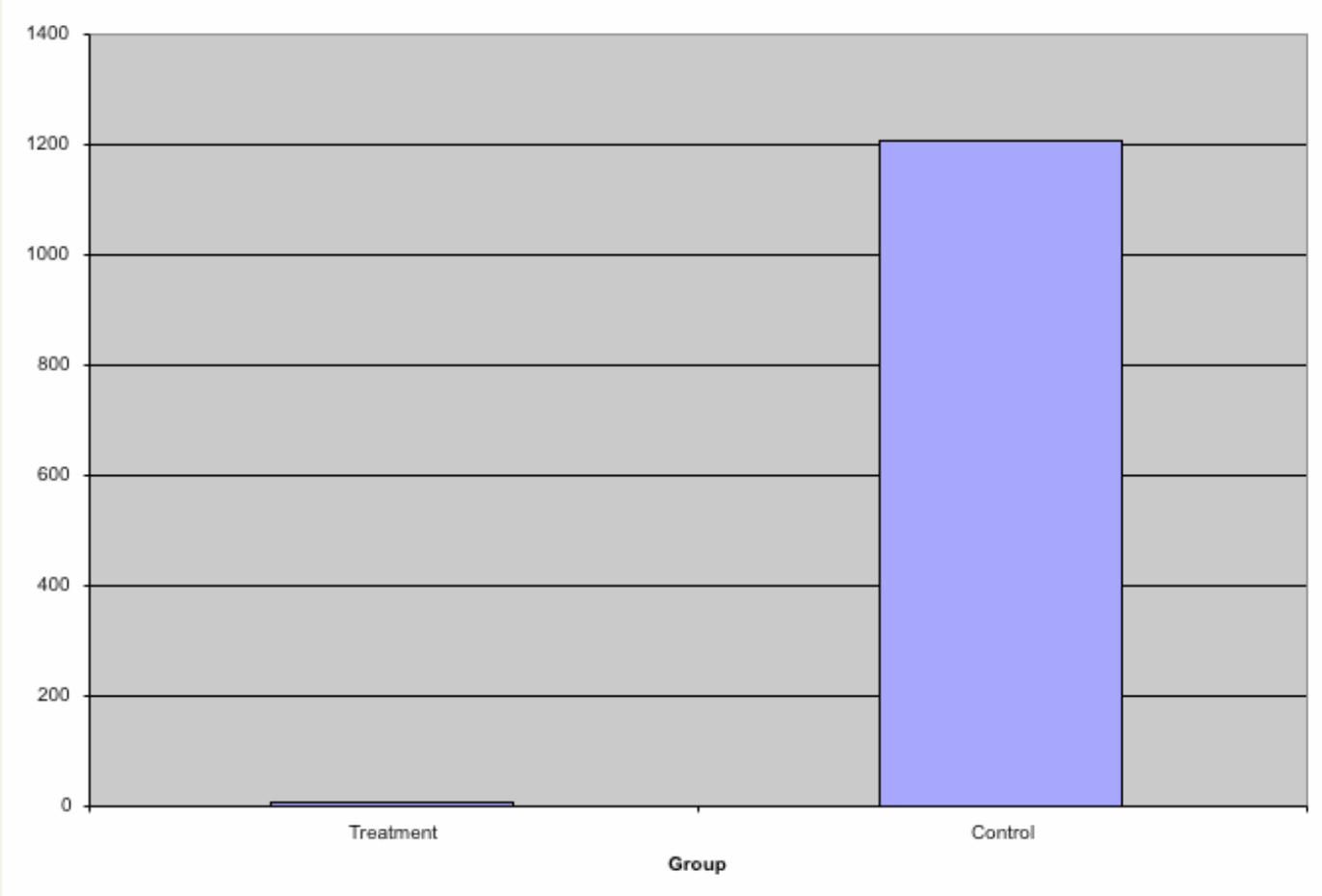
Example 3



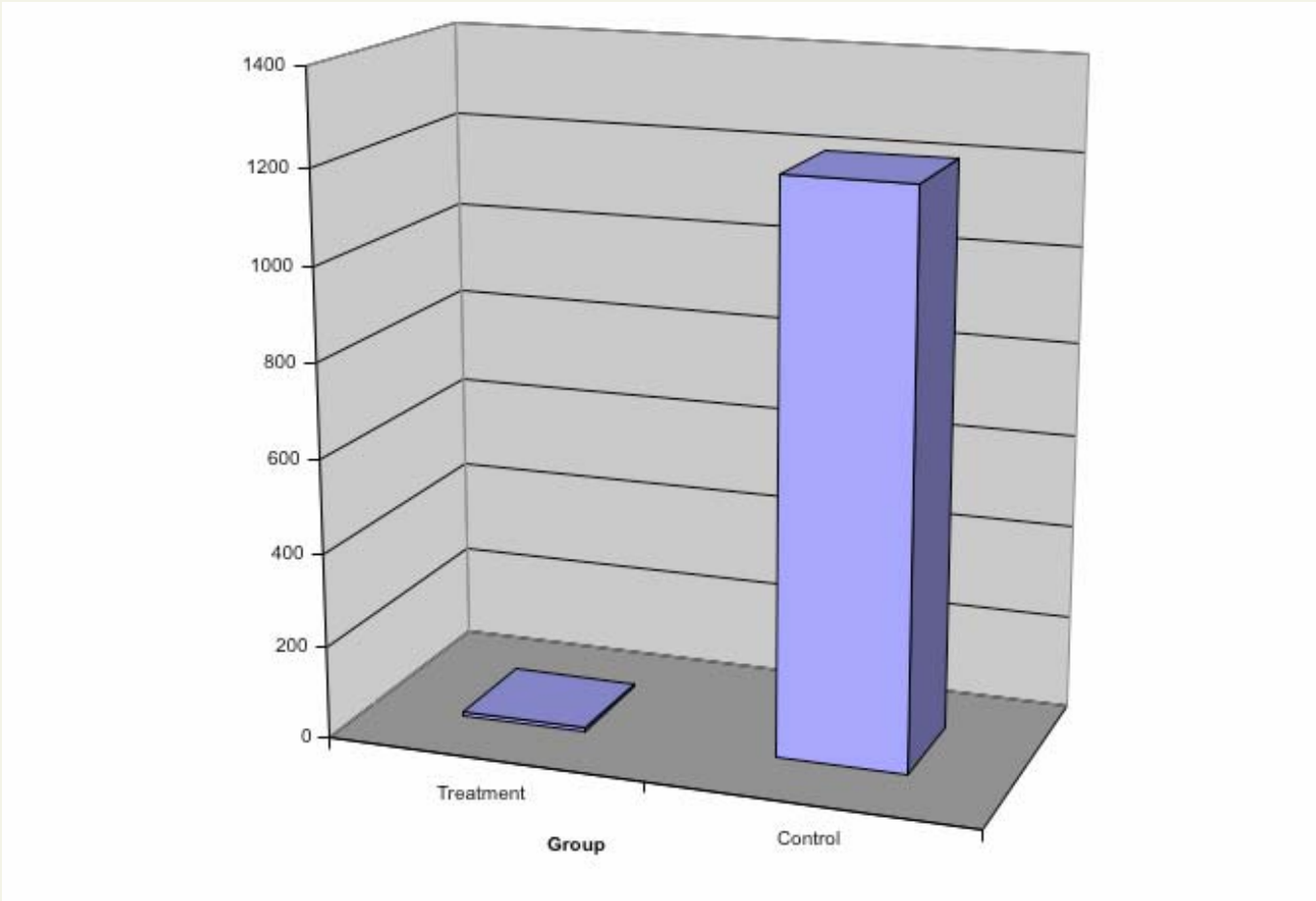
Example 3



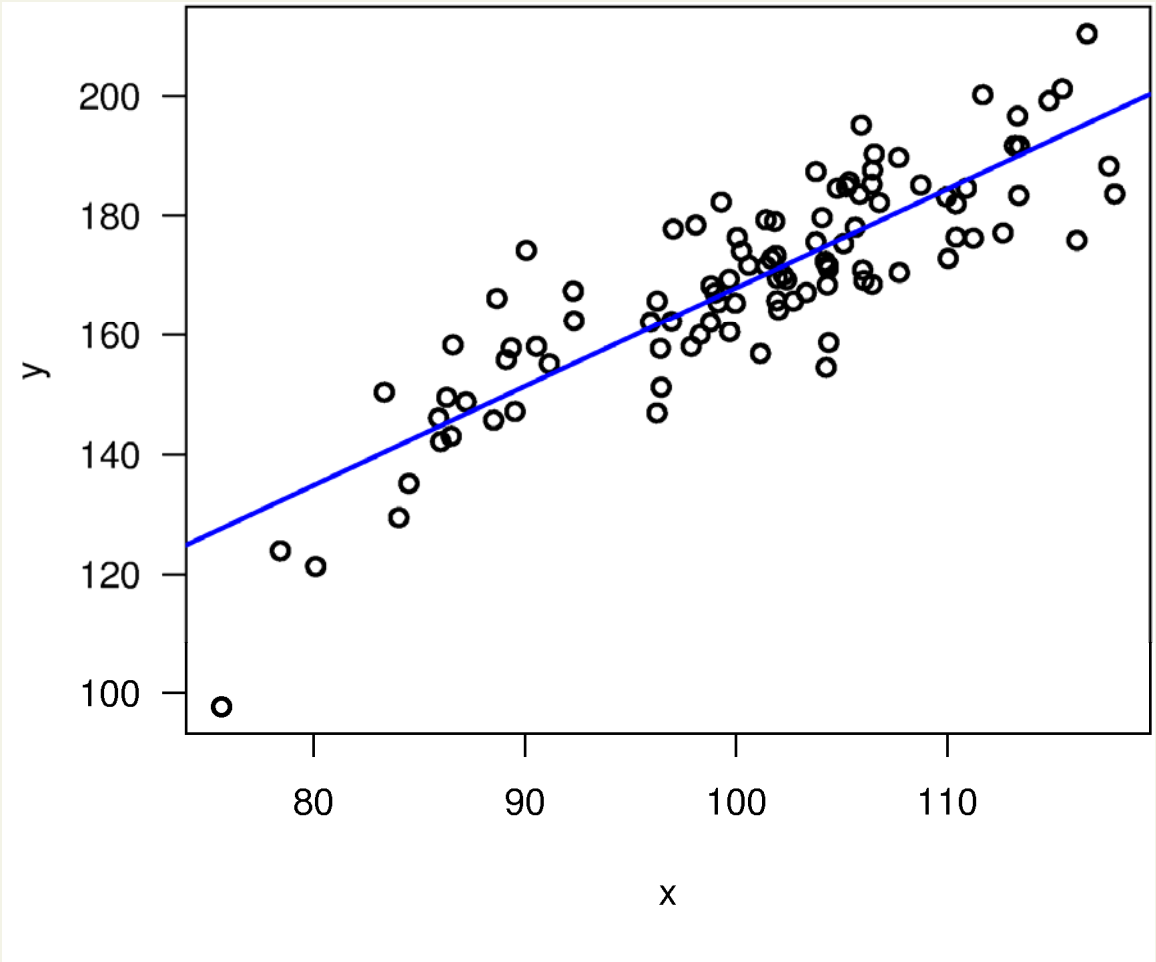
Example 3



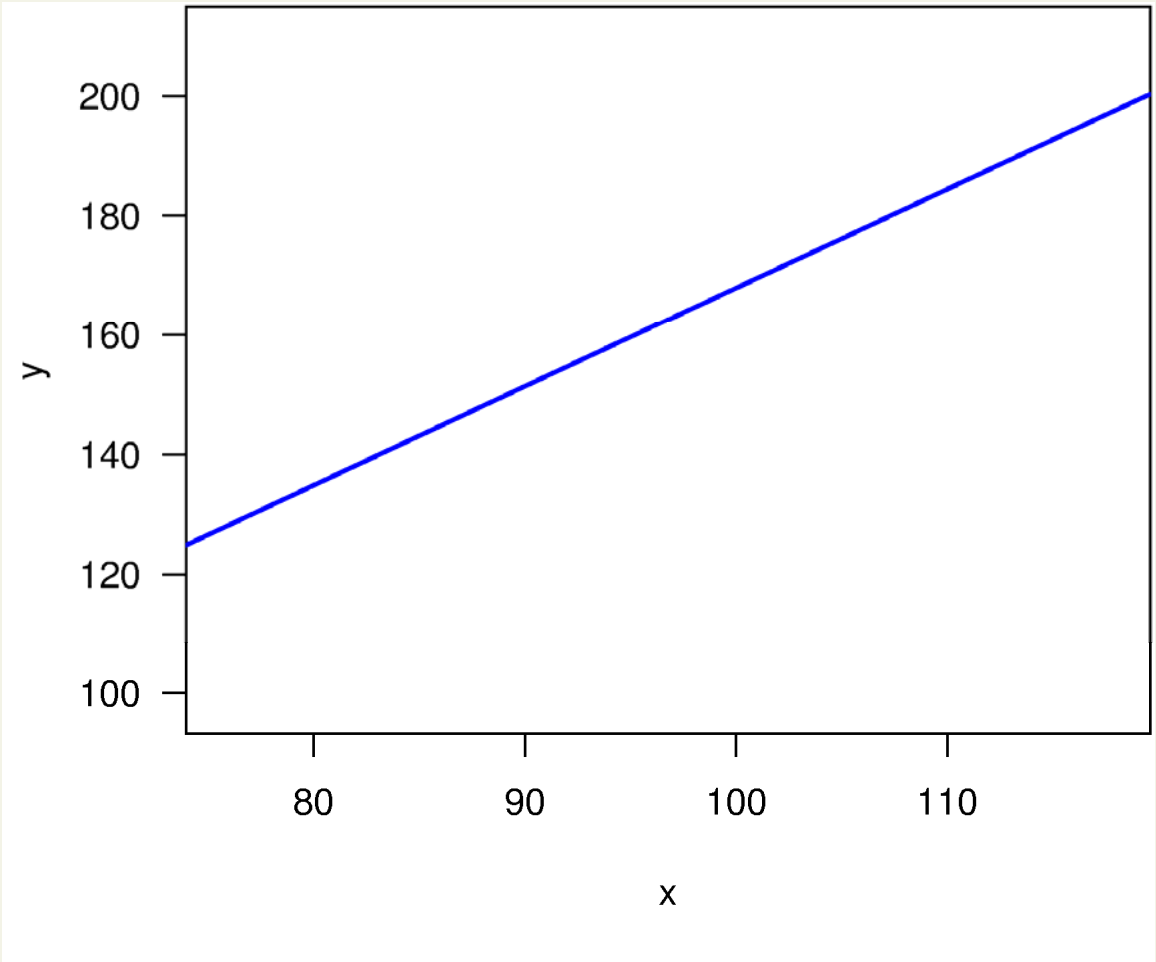
Example 3



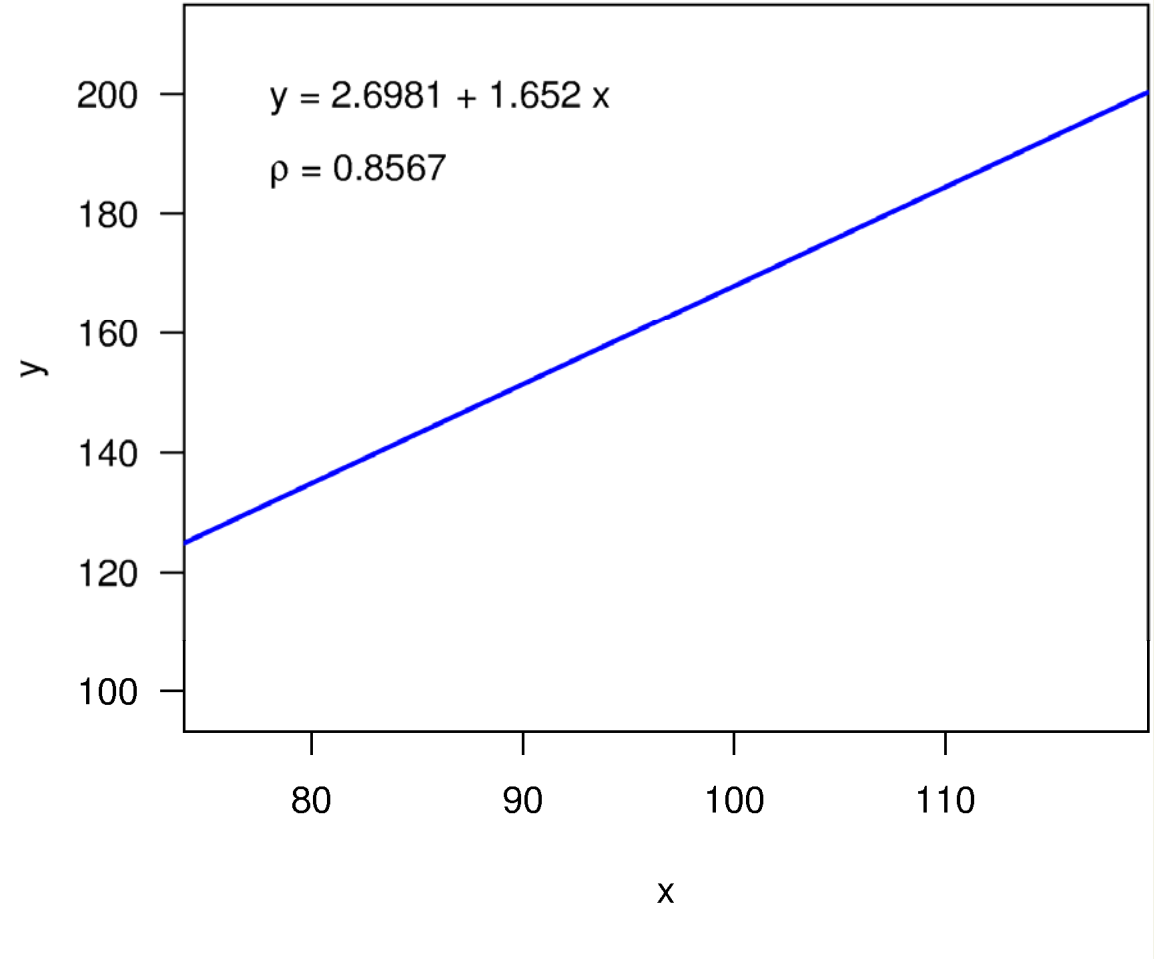
Example 4



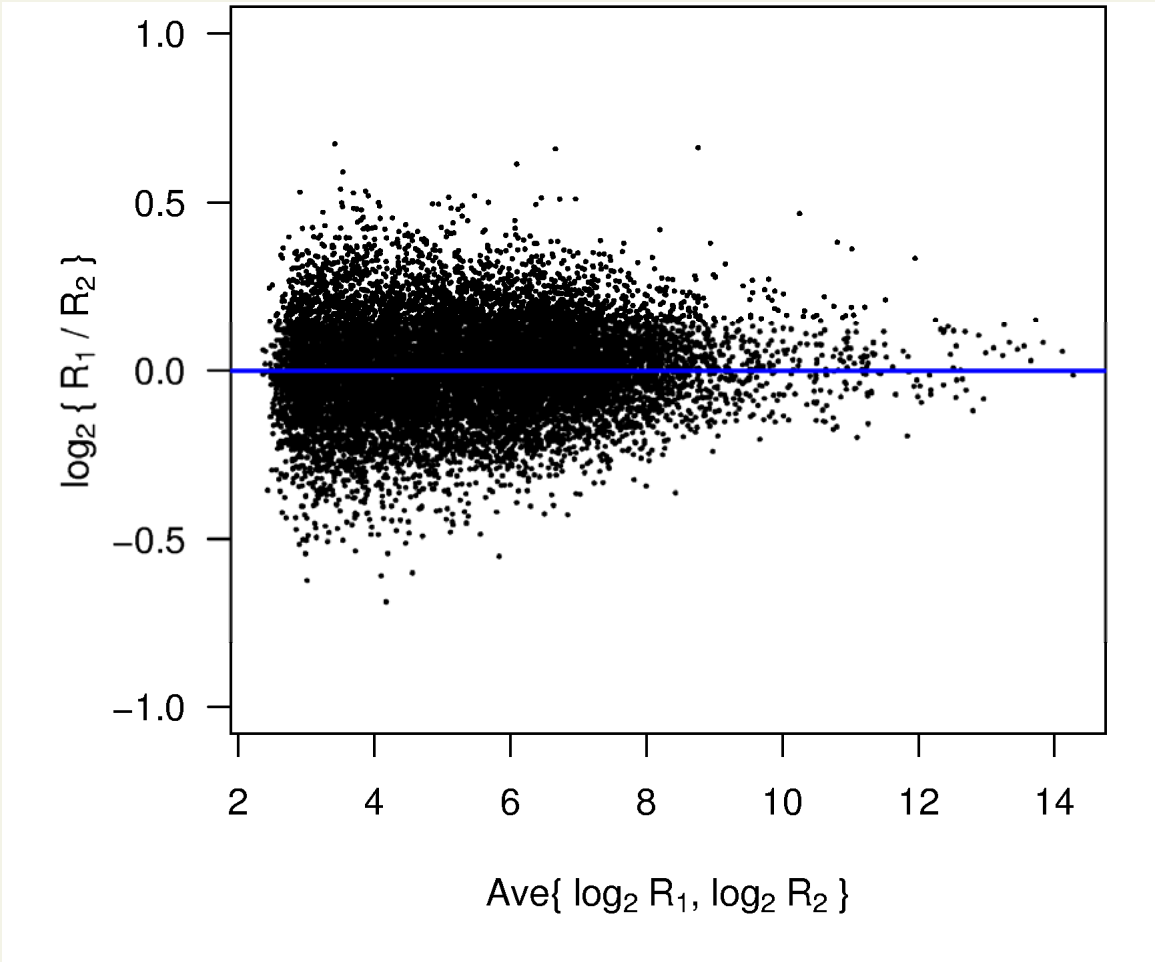
Example 4



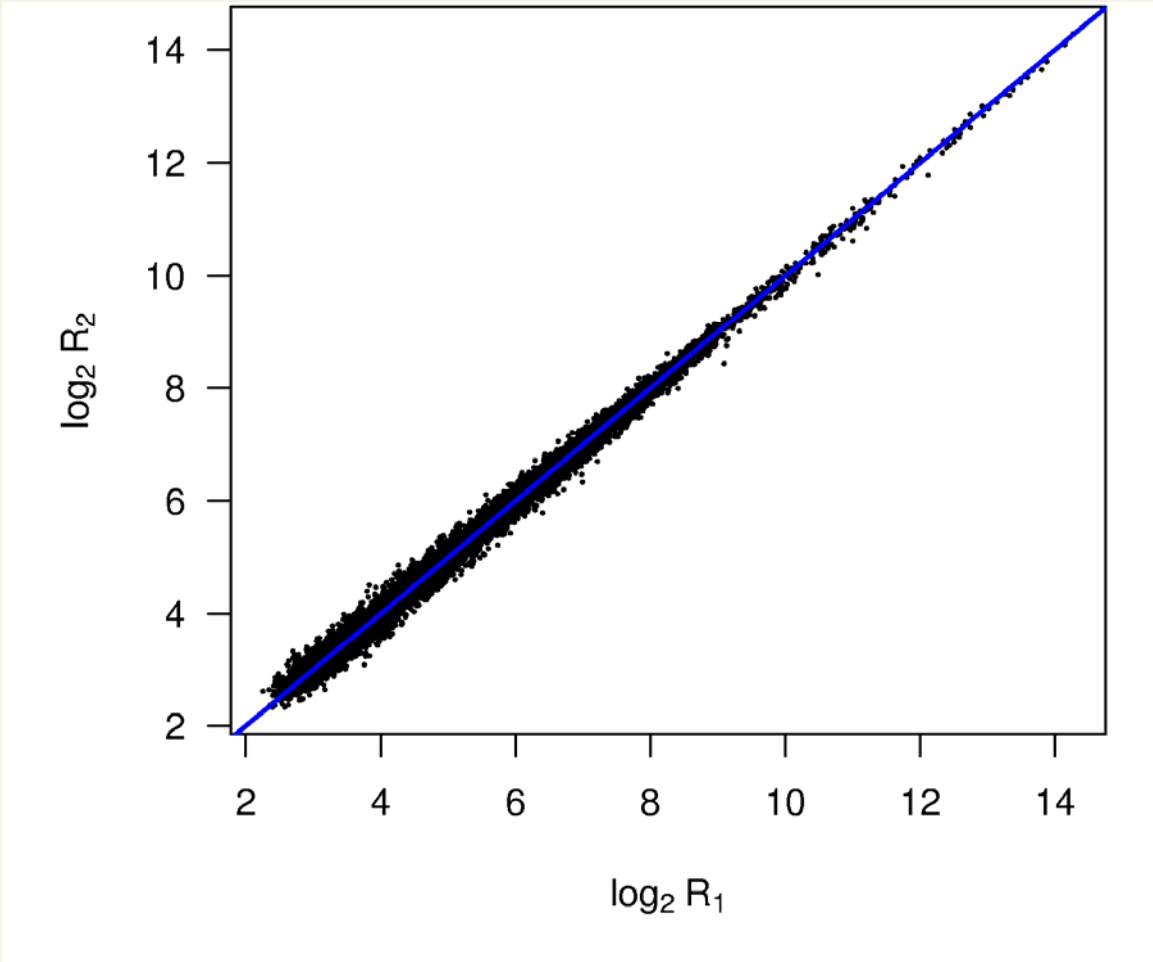
Example 4



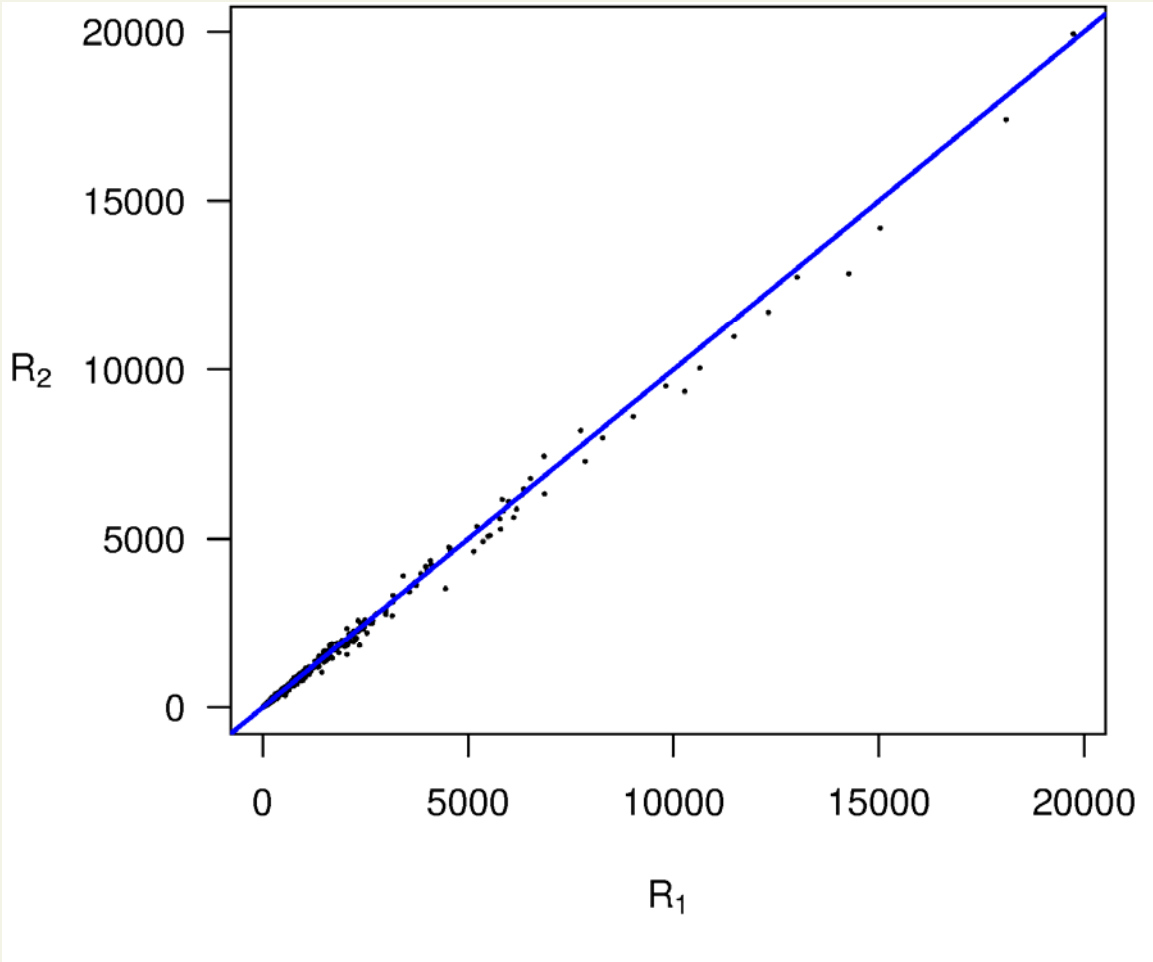
Example 5



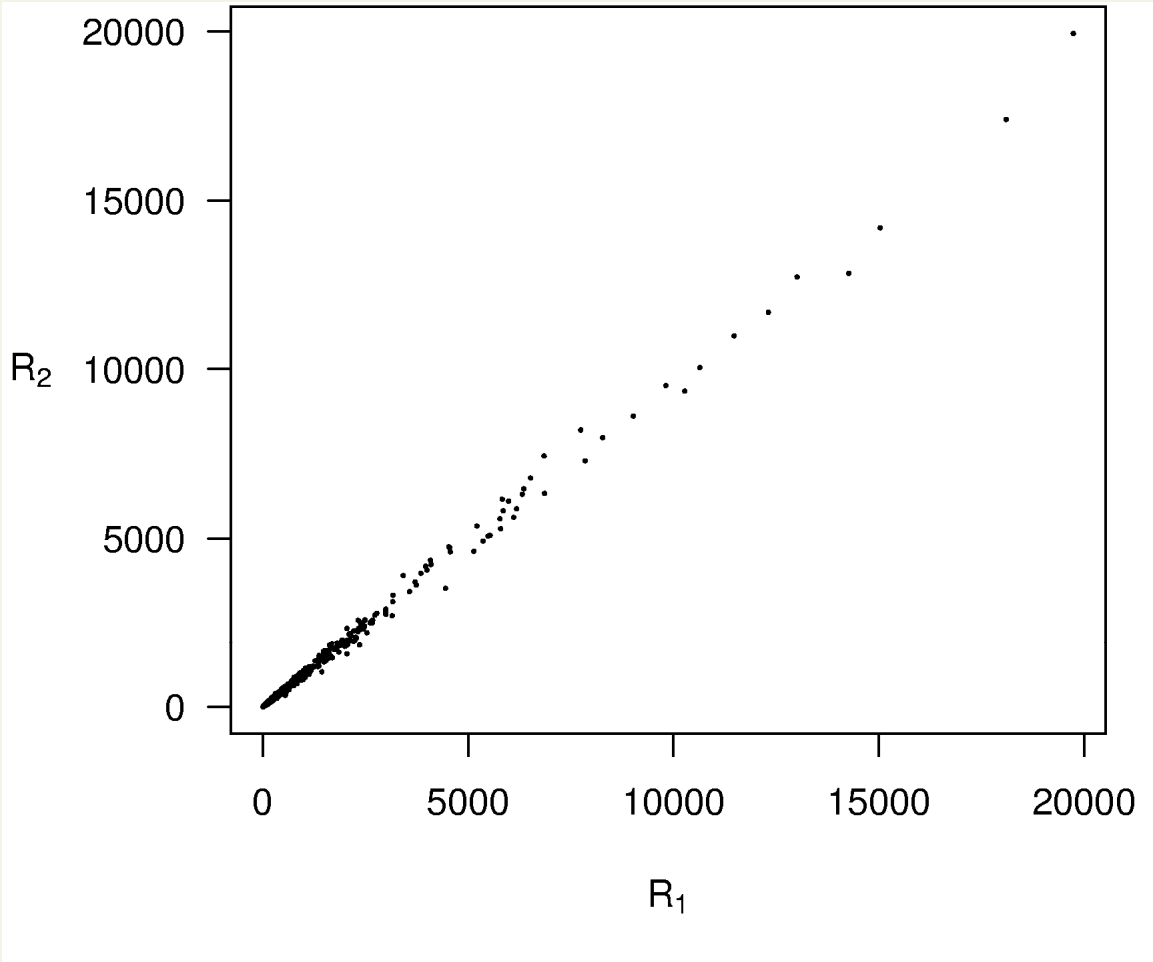
Example 5



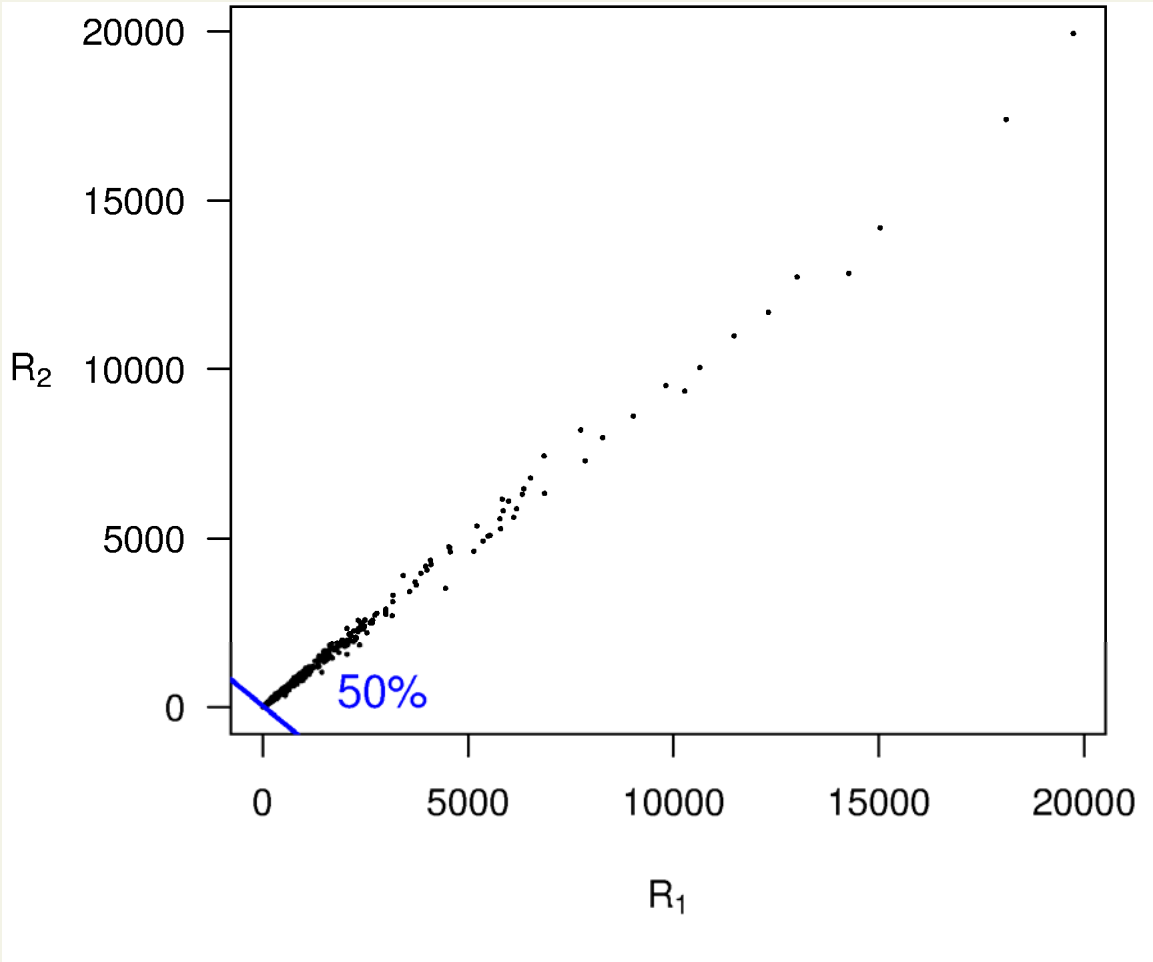
Example 5



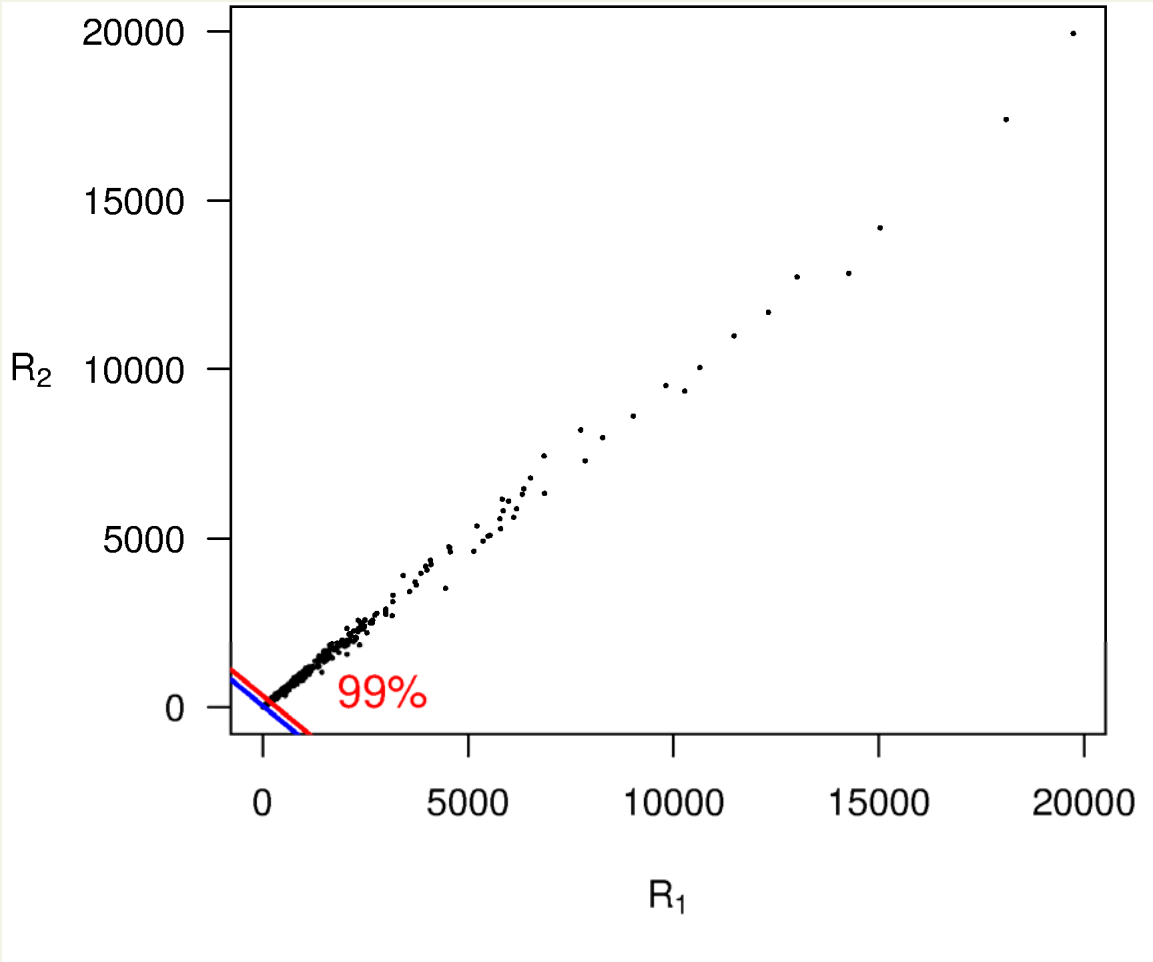
Example 5



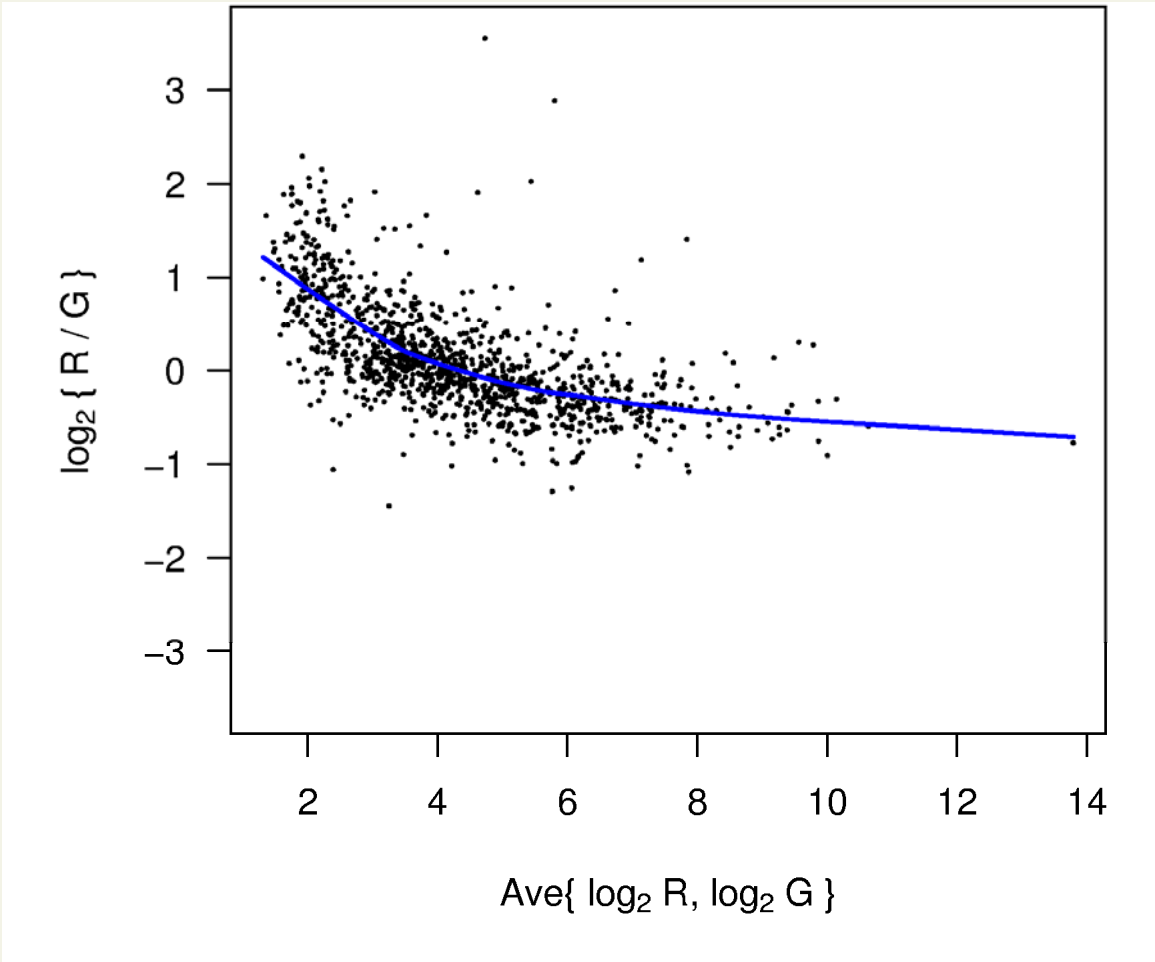
Example 5



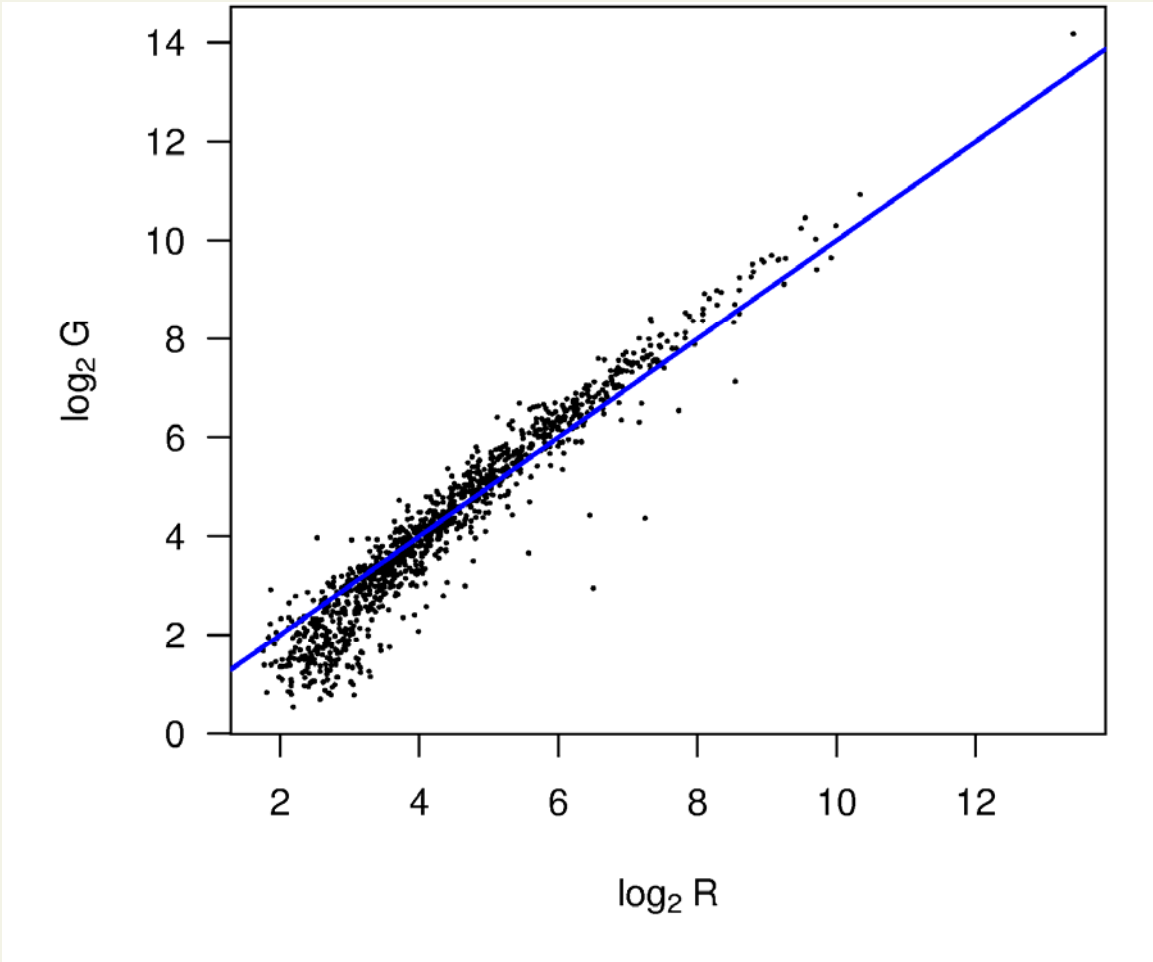
Example 5



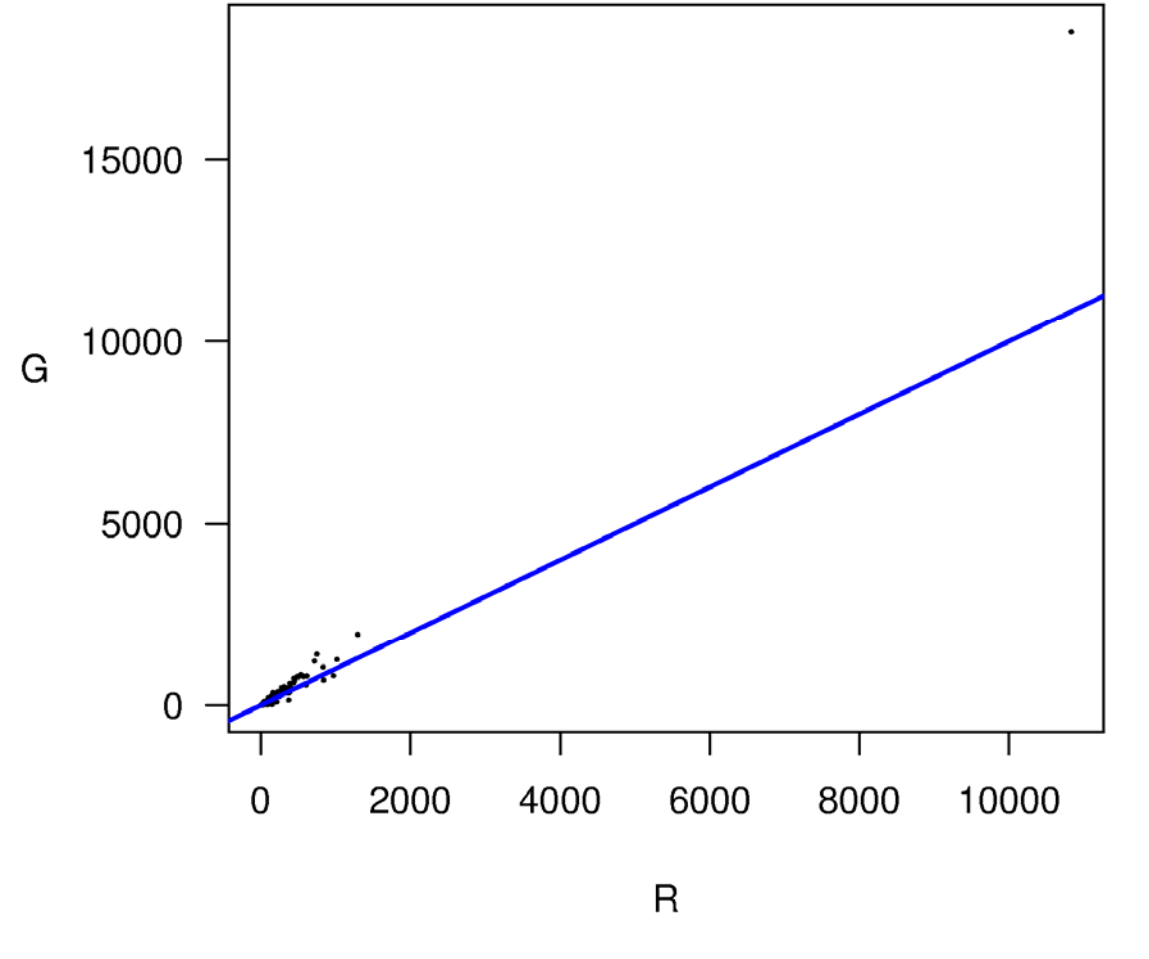
Example 6



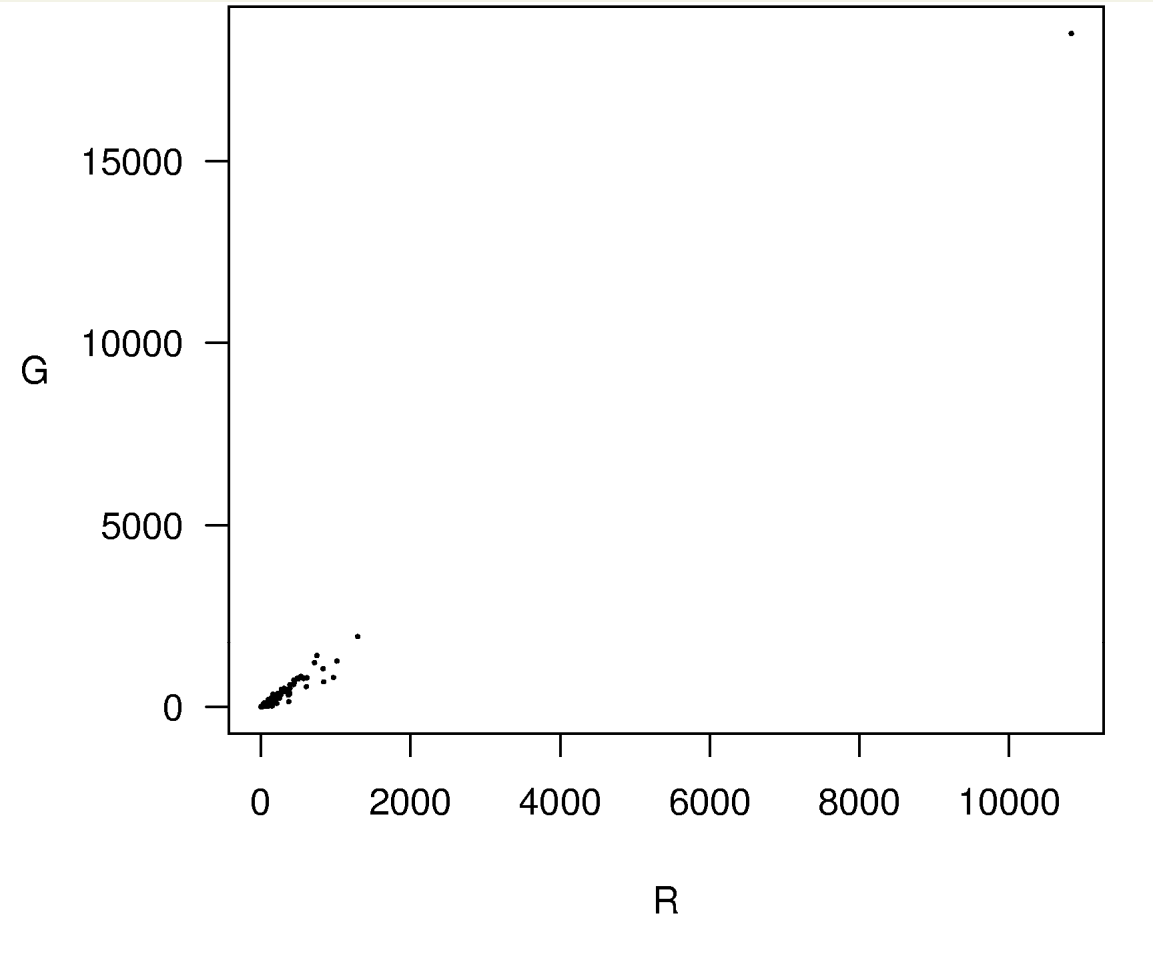
Example 6



Example 6



Example 6



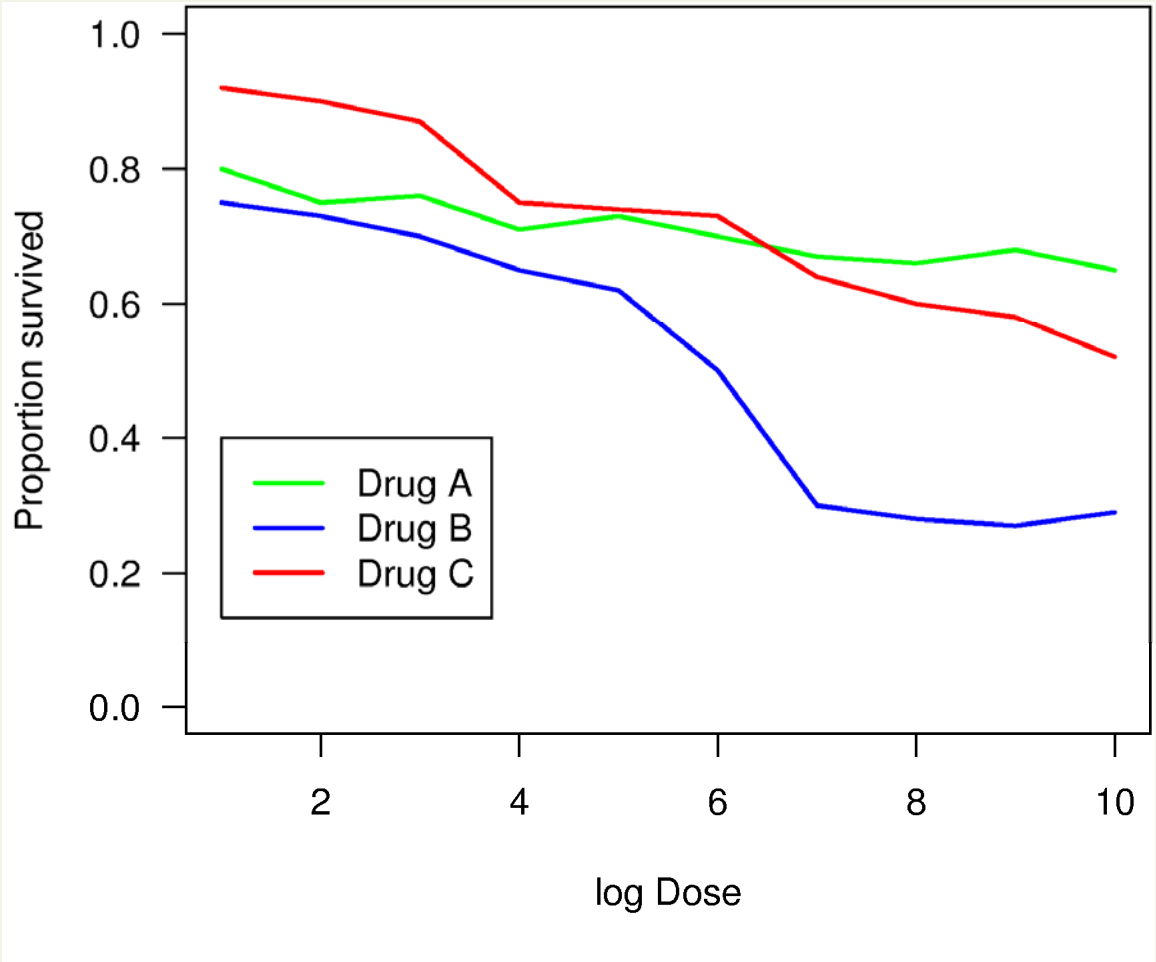
Example 7

N	$b/c = 10.0$		$b/c = 10.0$		$b/c = 100.0$	
	$r^\star$	G	$r^\star$	G	$r^\star$	G
3	2	0.20	2	2.2	2	22
4	2	0.26	2	2.9	2	29
5	2	0.32	3	3.5	3	36
6	3	0.38	3	4.2	3	43
7	3	0.45	3	4.9	3	49
8	3	0.51	4	5.6	4	56
9	3	0.57	4	6.3	4	63
10	4	0.63	4	6.9	4	70

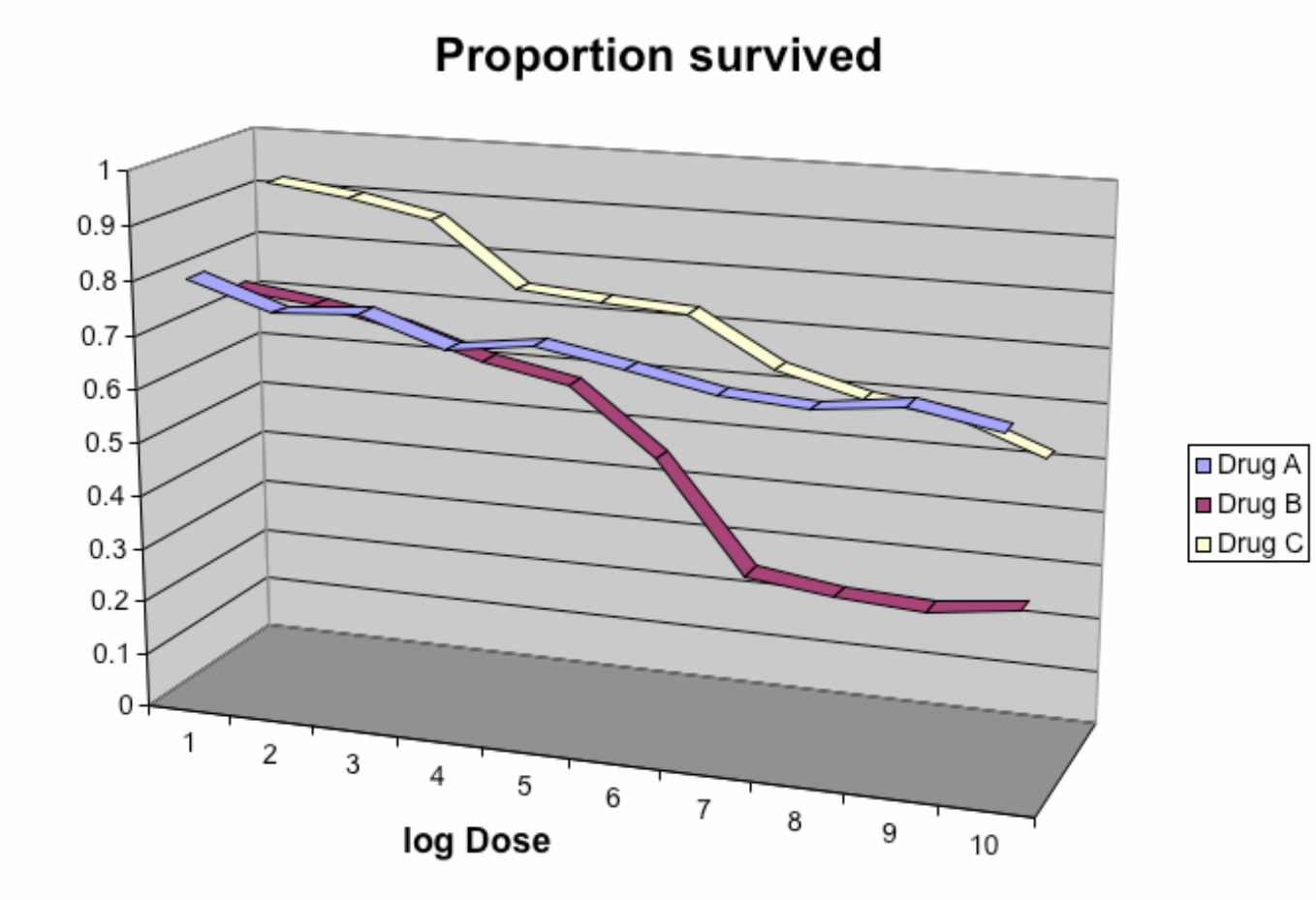
Example 7

	$b/c = 10.0$		$b/c = 10.0$		$b/c = 100.0$	
N	$r^\star$	G	$r^\star$	G	$r^\star$	G
3	2	0.2	2	2.225	2	22.47499
4	2	0.26333	2	2.88833	2	29.13832
5	2	0.32333	3	3.54167	3	35.79166
6	3	0.38267	3	4.23767	3	42.78764
7	3	0.446	3	4.901	3	49.45097
8	3	0.50743	4	5.5765	4	56.33005
9	3	0.56743	4	6.26025	4	63.20129
10	4	0.62948	4	6.92358	4	69.86462

Example 8


1.0
0.8
0.6
0.4
0.2
0.0
Proportion survived
Drug A
Drug B
Drug C
2
4
6
8
10
log Dose

Example 8


Proportion survived
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
log Dose
Drug A
Drug B
Drug C

Displaying data well

- Be accurate and clear.
- Let the data speak.
  - Show as much information as possible, taking care not to obscure the message.
- Science not sales.
  - Avoid unnecessary frills — esp. gratuitous 3d.
- In tables, every digit should be meaningful. Don’t drop ending 0’s.

Further reading

- ER Tufte (1983) The visual display of quantitative information. Graphics Press.
- ER Tufte (1990) Envisioning information. Graphics Press.
- ER Tufte (1997) Visual explanations. Graphics Press.
- WS Cleveland (1993) Visualizing data. Hobart Press.
- WS Cleveland (1994) The elements of graphing data. CRC Press.