Creating effective figures and tables

Karl W Broman

Biostatistics & Medical Informatics
University of Wisconsin – Madison

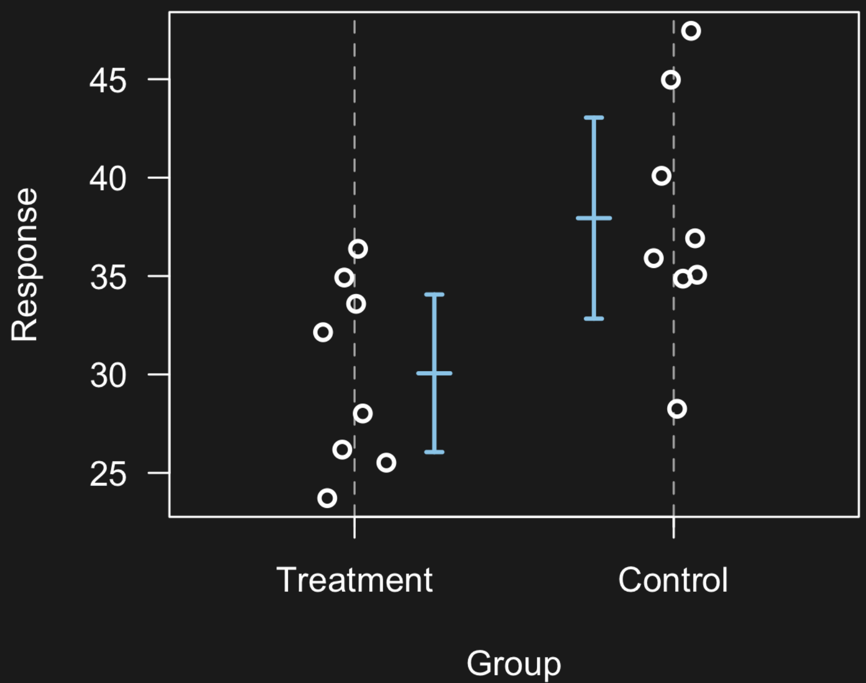
kbroman.org
github.com/kbroman
@kwbroman

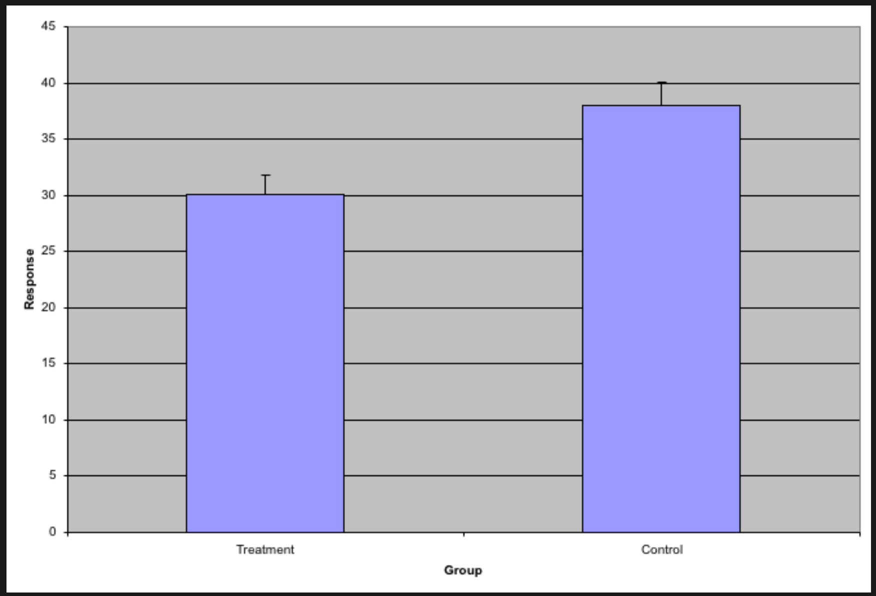
Slides: bit.ly/CDHA2016

Displaying data well

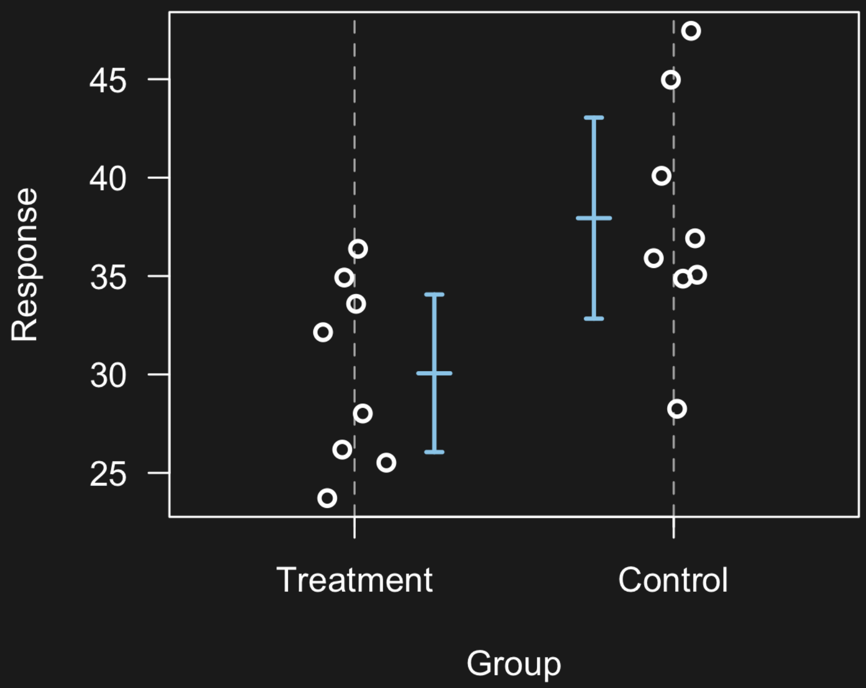
- Be accurate and clear.
- Let the data speak.
  - Show as much information as possible, taking care not to obscure the message.
- Science not sales.
  - Avoid unnecessary frills (esp. gratuitous 3d).
- In tables, every digit should be meaningful. Don't drop ending 0's.

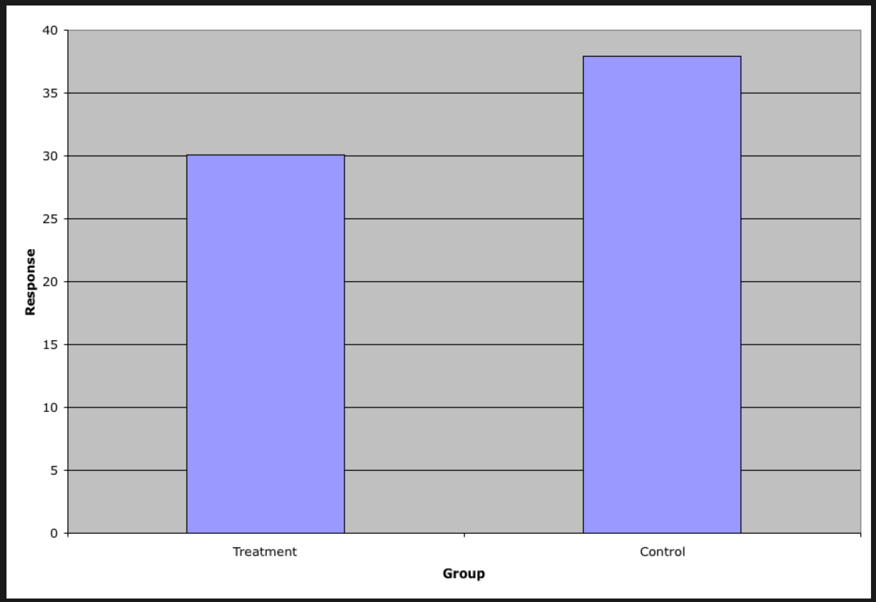
Show the data



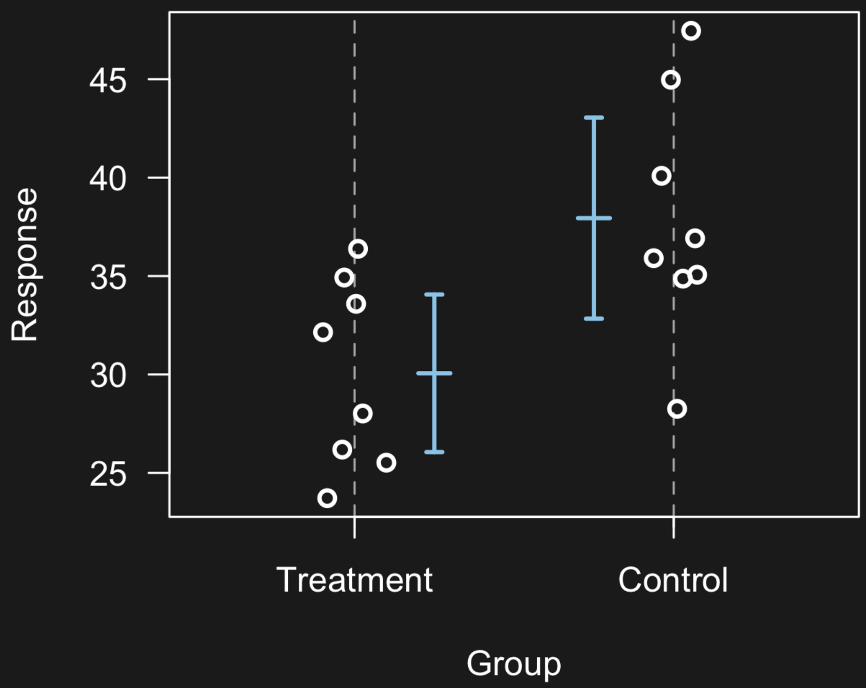


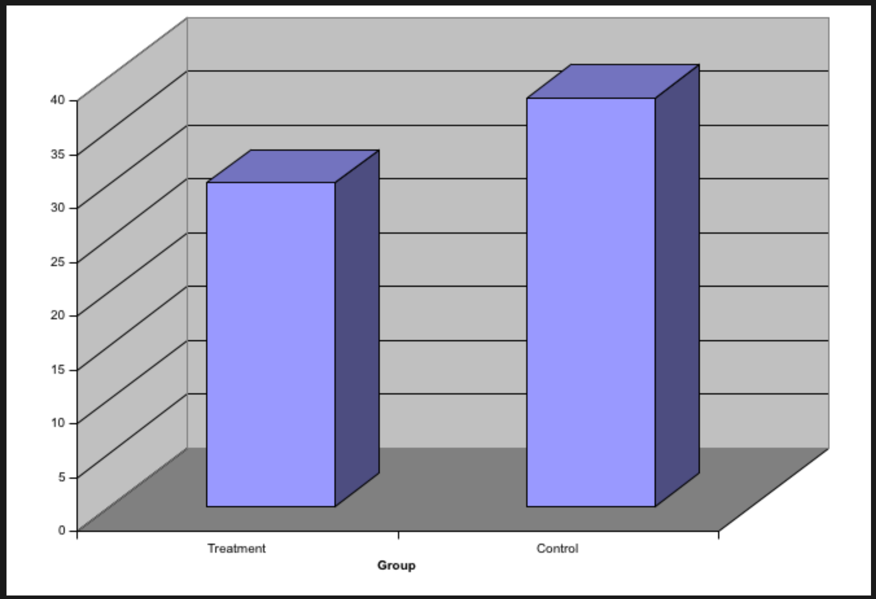
Show the data



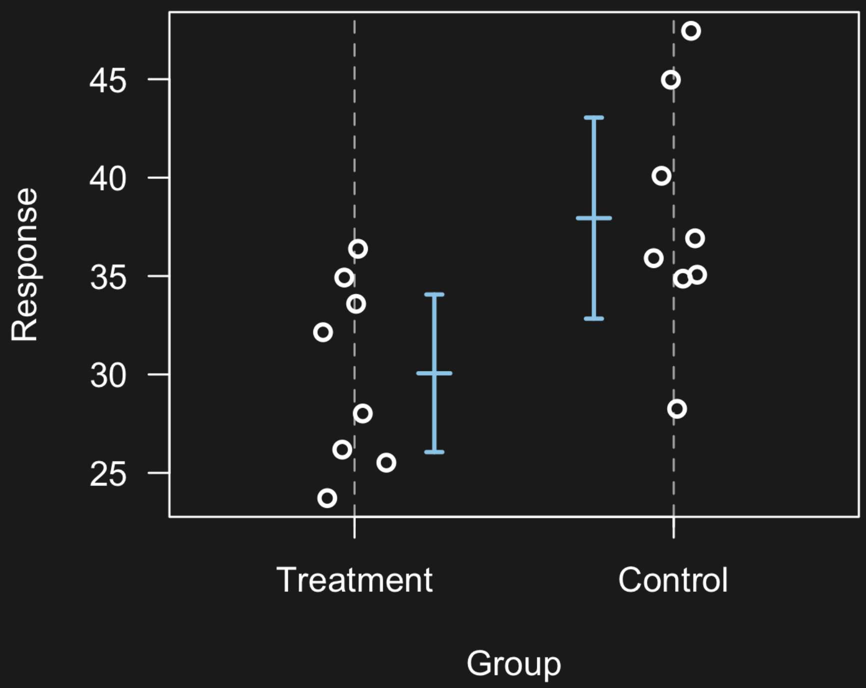


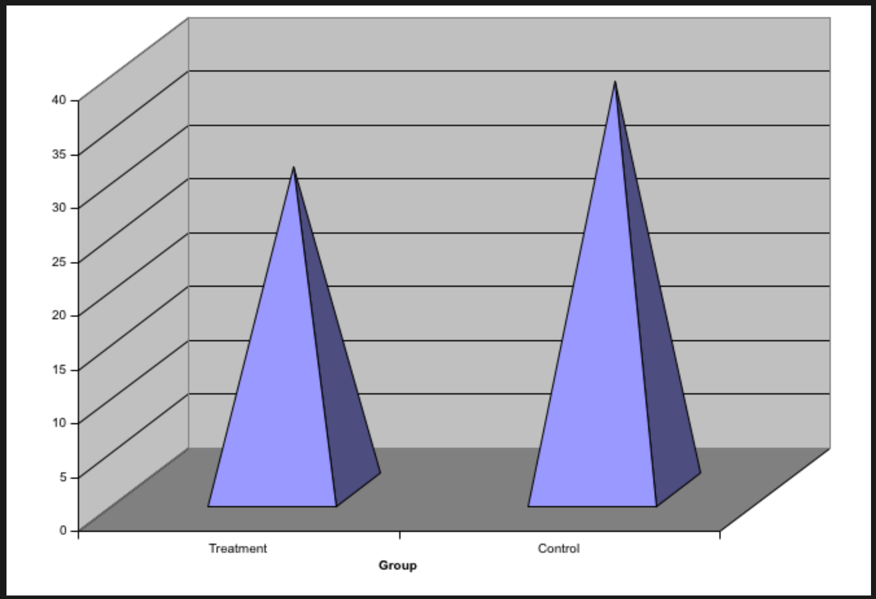
Show the data



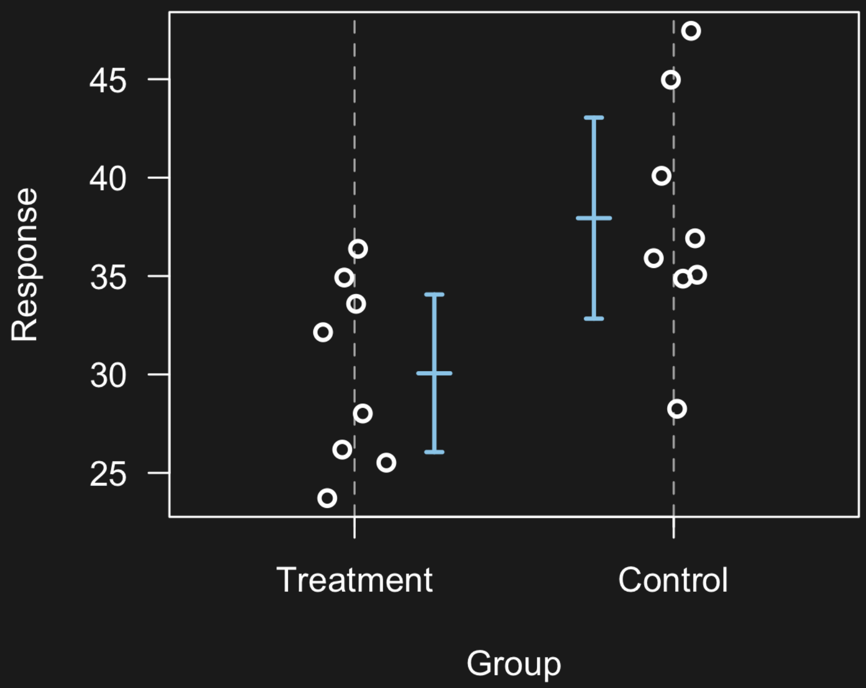


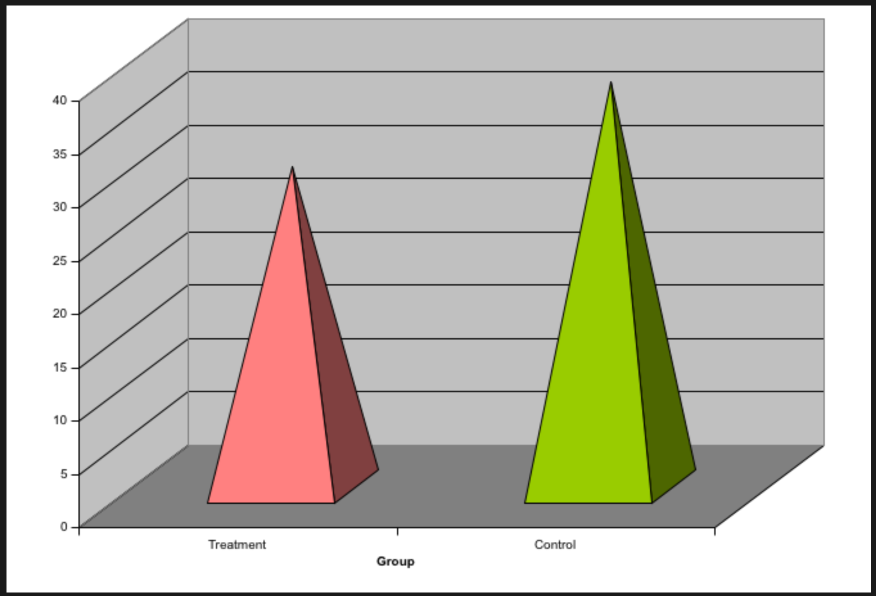
Show the data



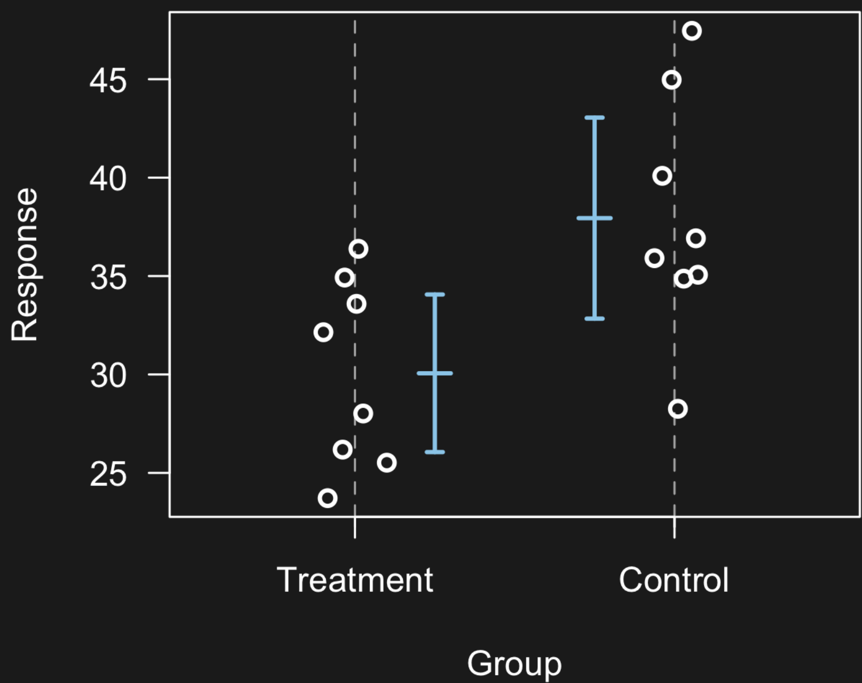


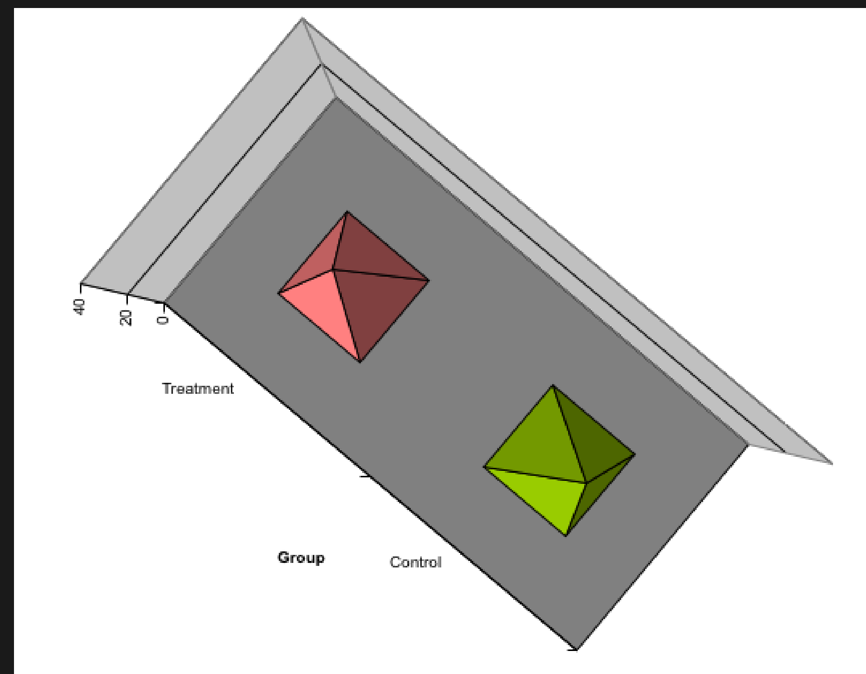
Show the data



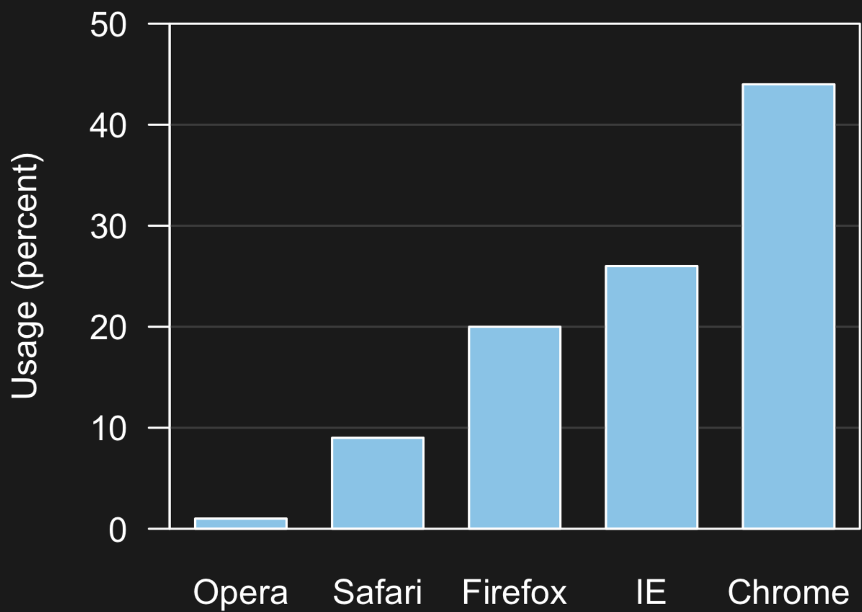


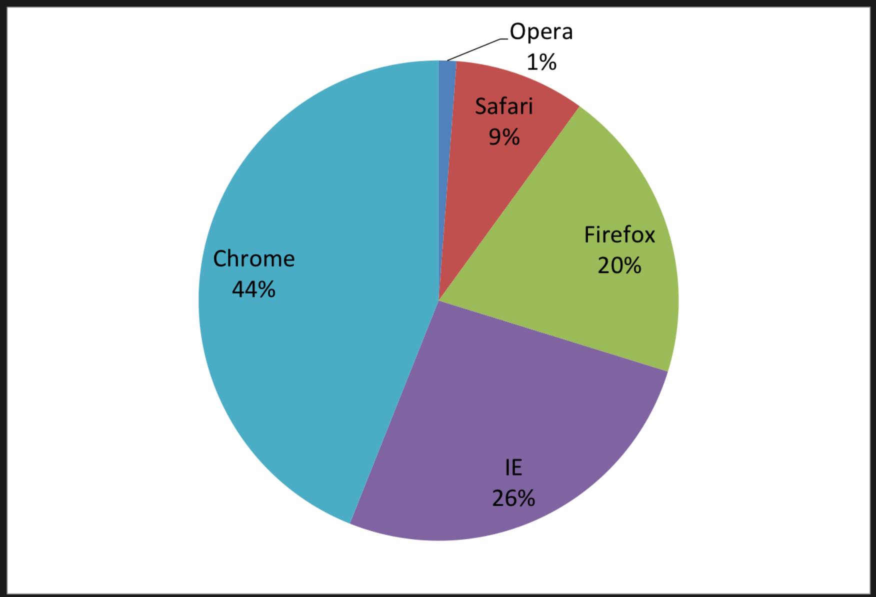
Show the data



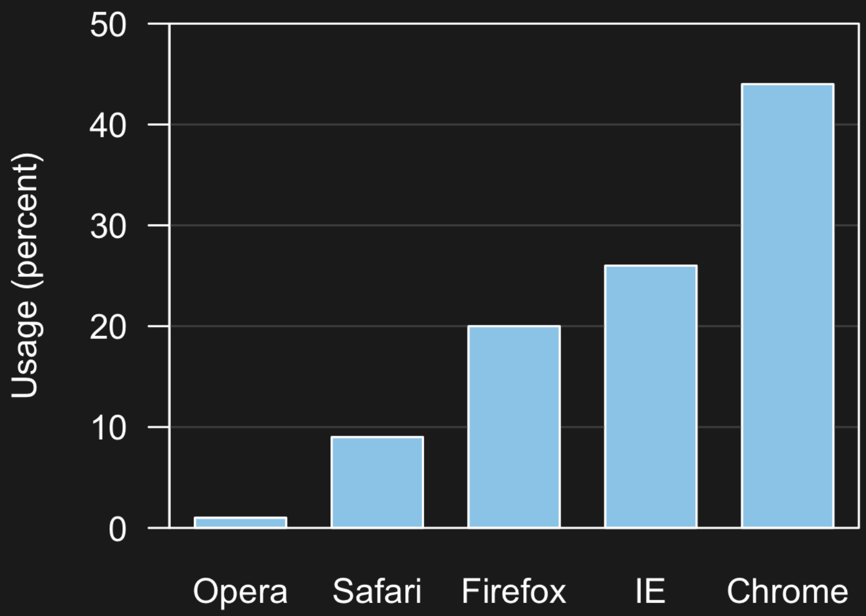


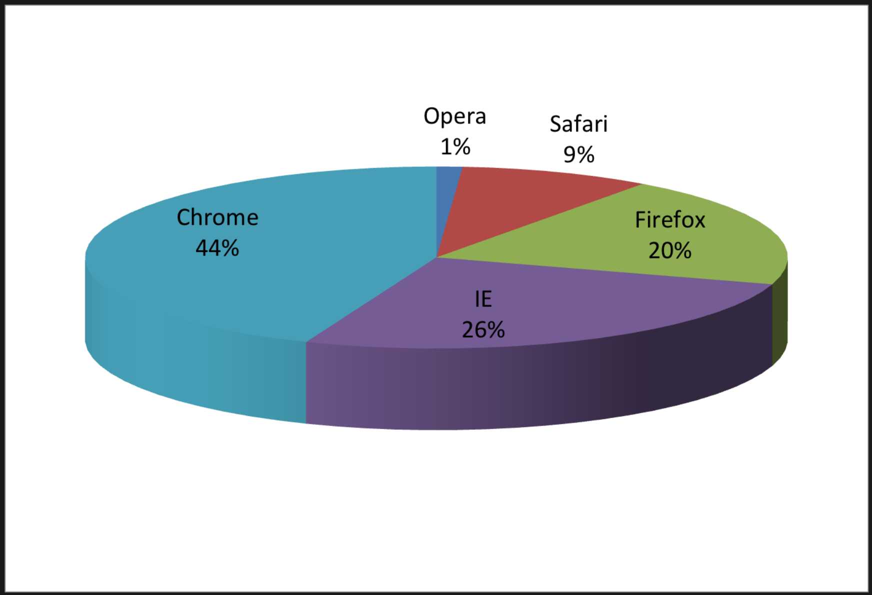
Avoid pie charts



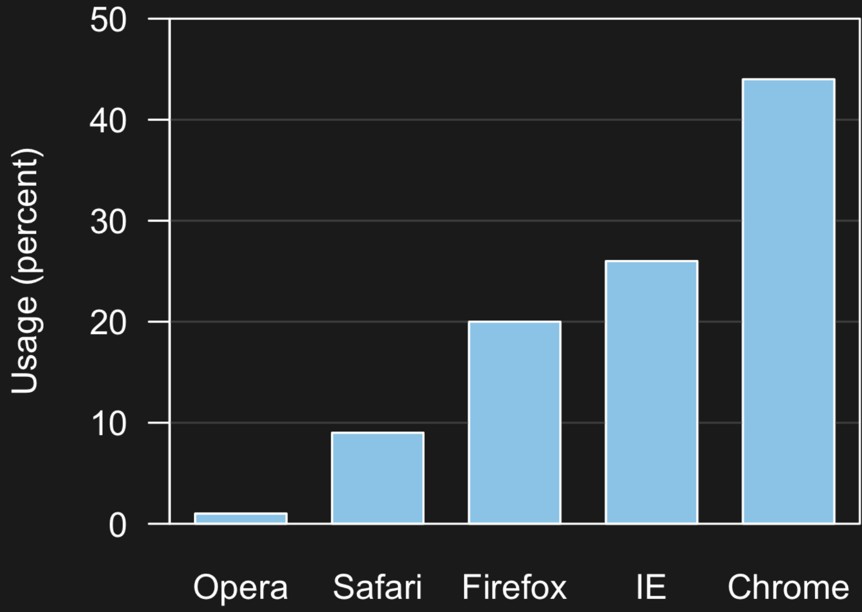


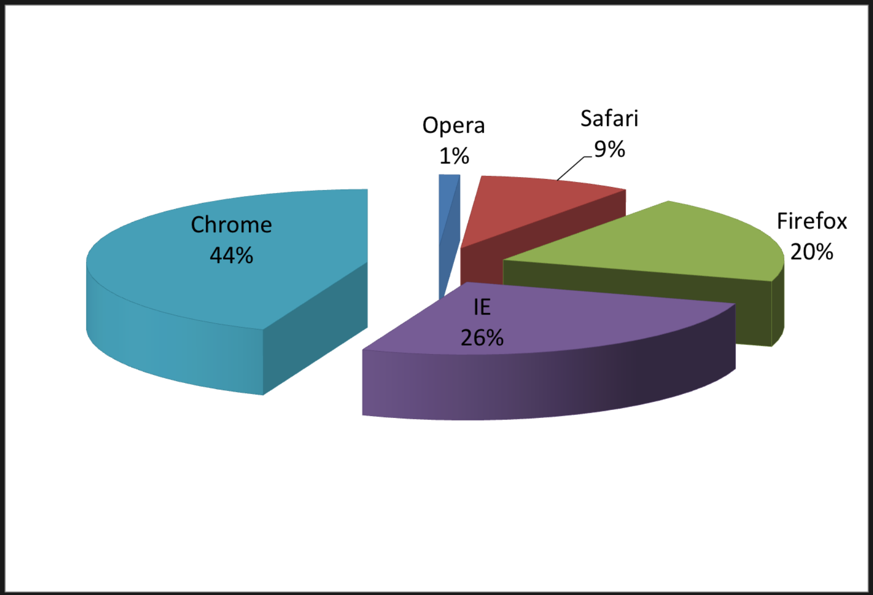
Avoid pie charts



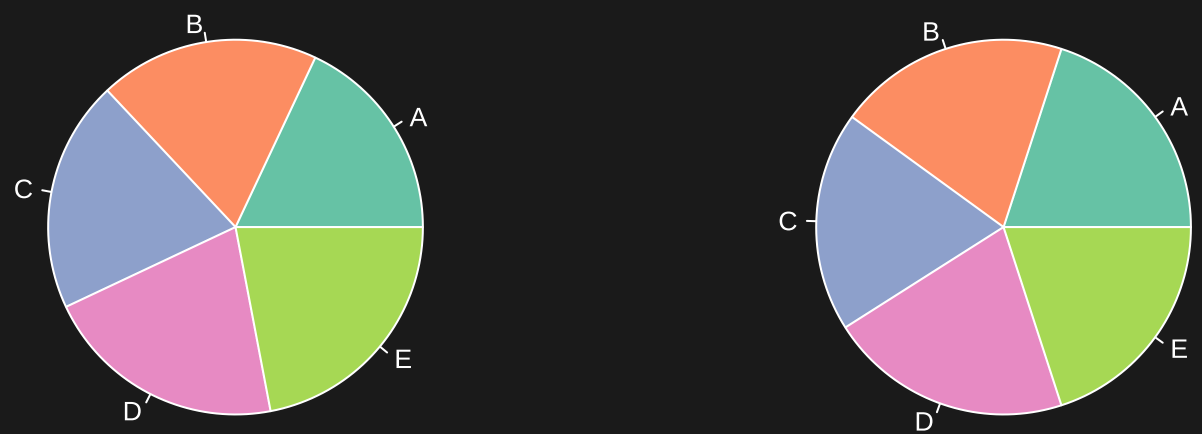


Avoid pie charts



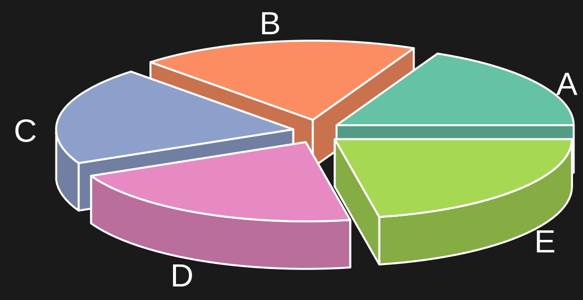


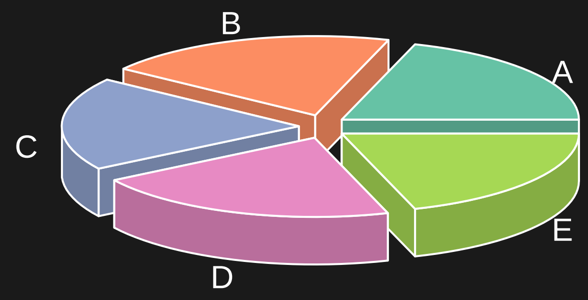
Avoid pie charts



via `@MonaChalabi` (`bit.ly/pie_vs_barchart`)

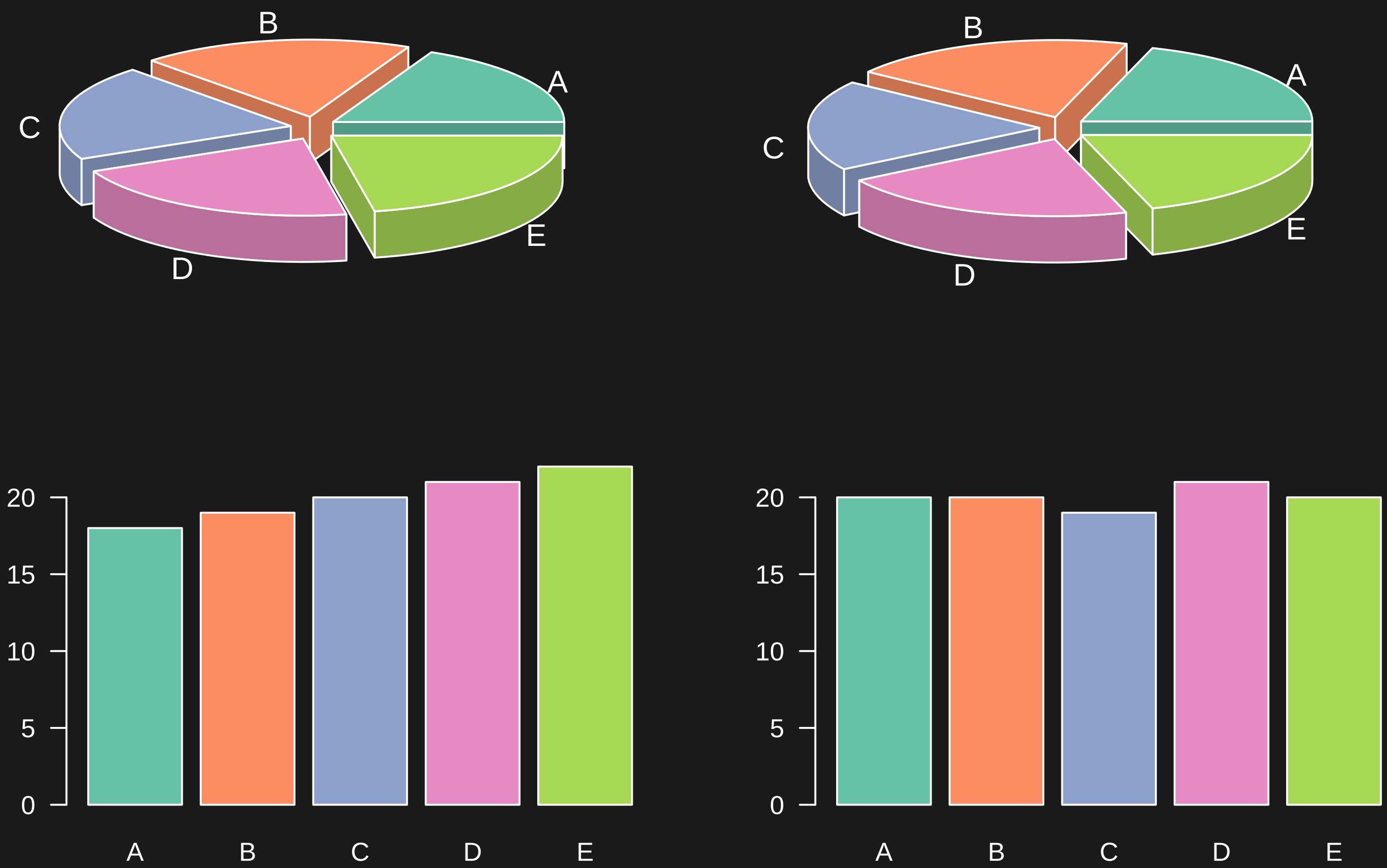
Avoid pie charts





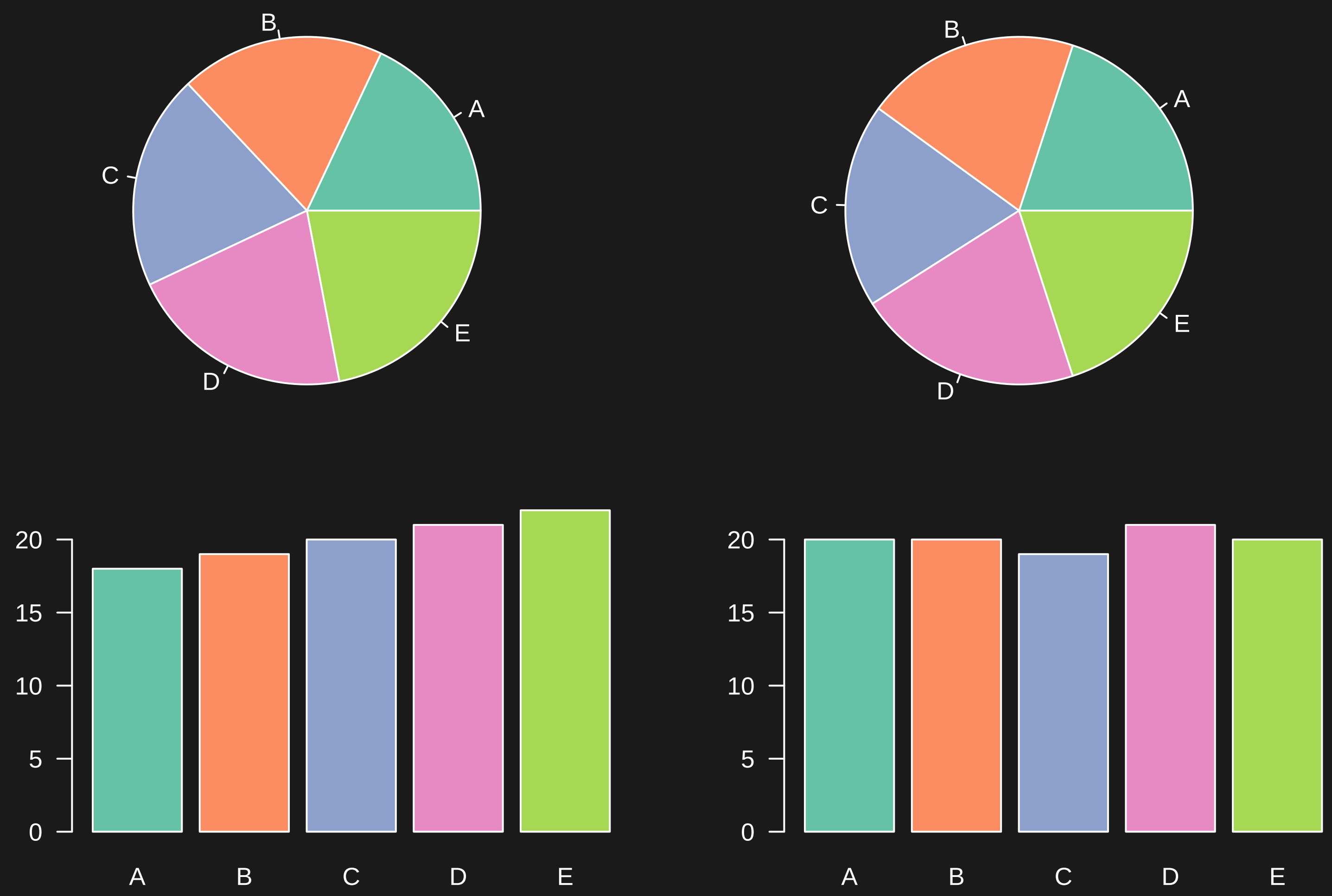
via @MonaChalabi (bit.ly/pie_vs_barchart)

Avoid pie charts



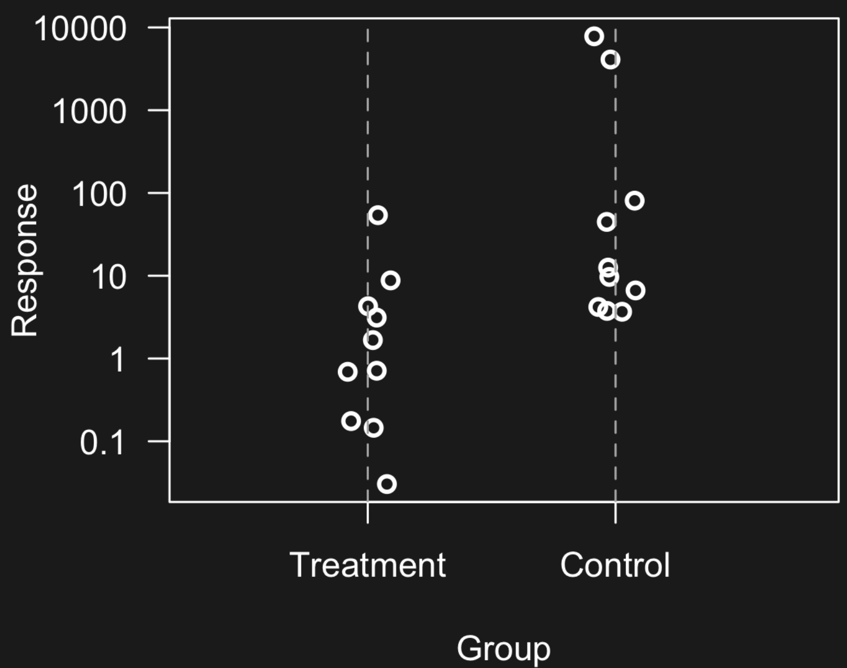
via @MonaChalabi (bit.ly/pie_vs_barchart)

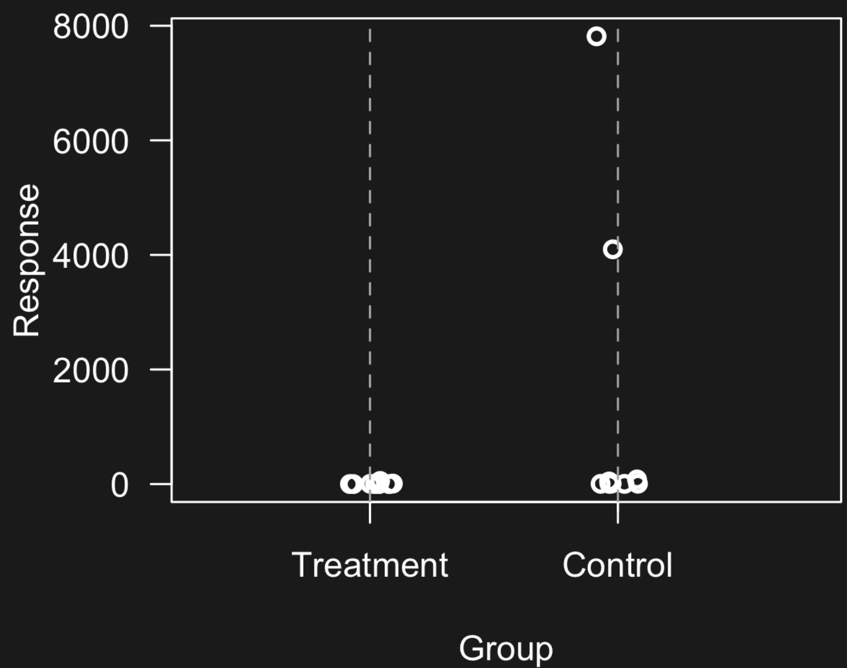
Avoid pie charts



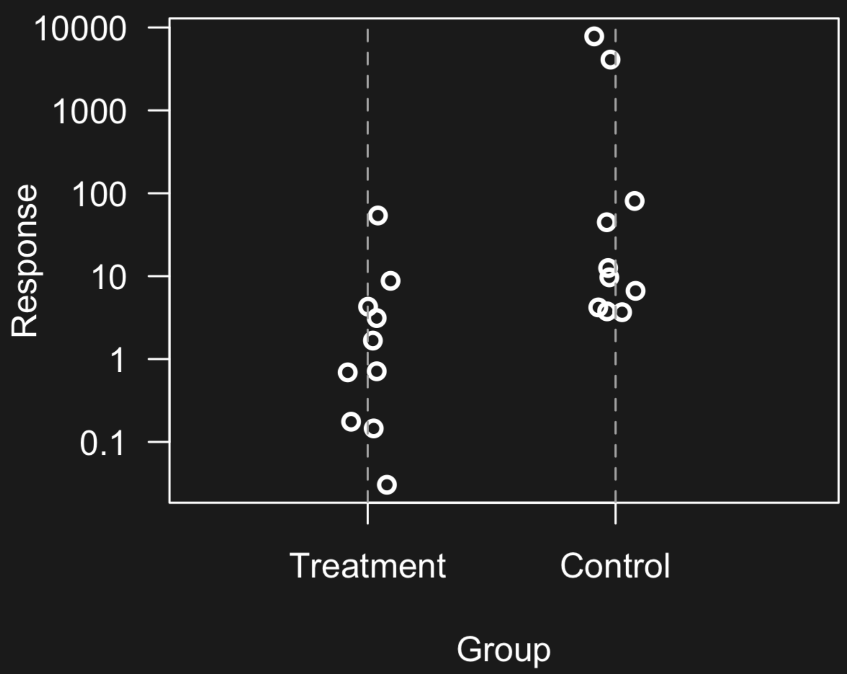
via @MonaChalabi (bit.ly/pie_vs_barchart)

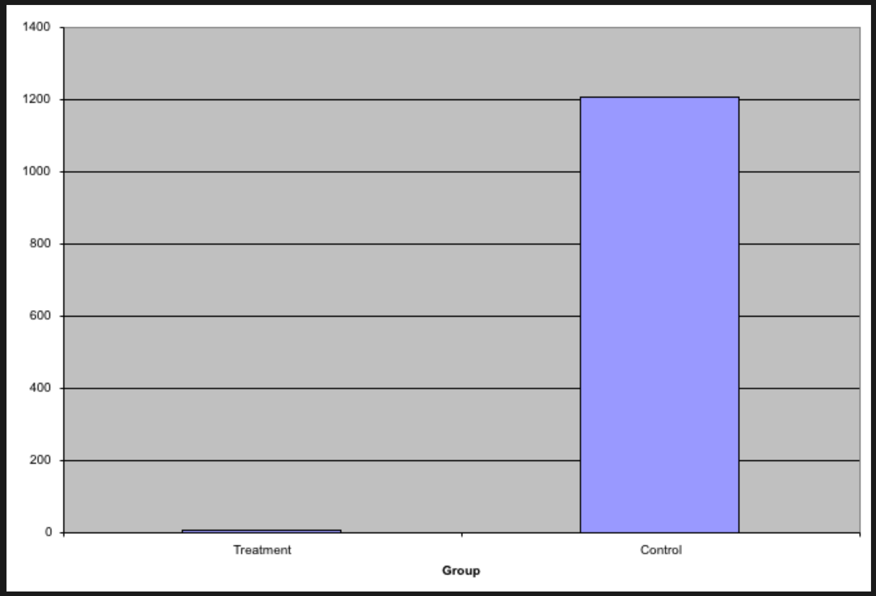
Consider logs



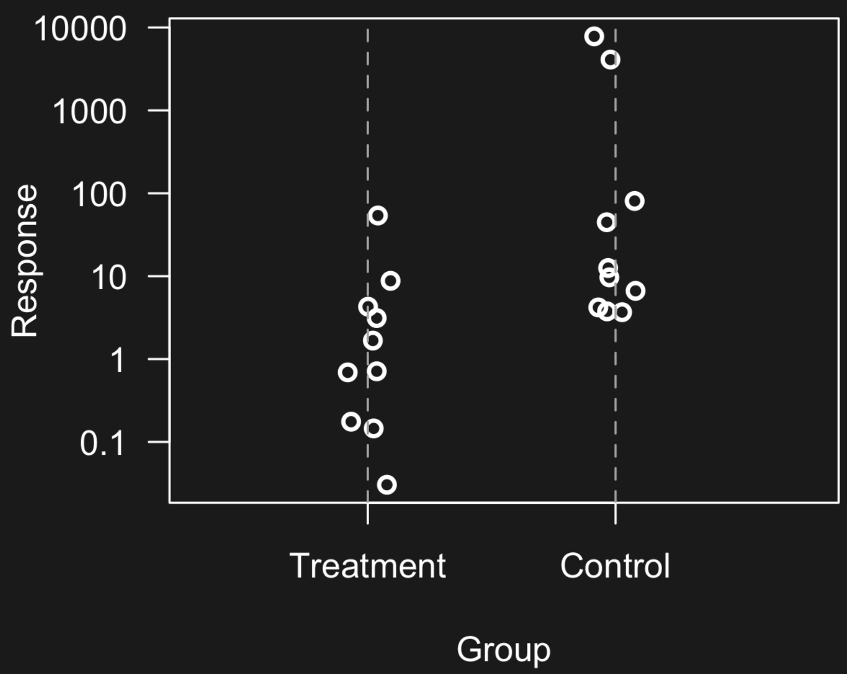


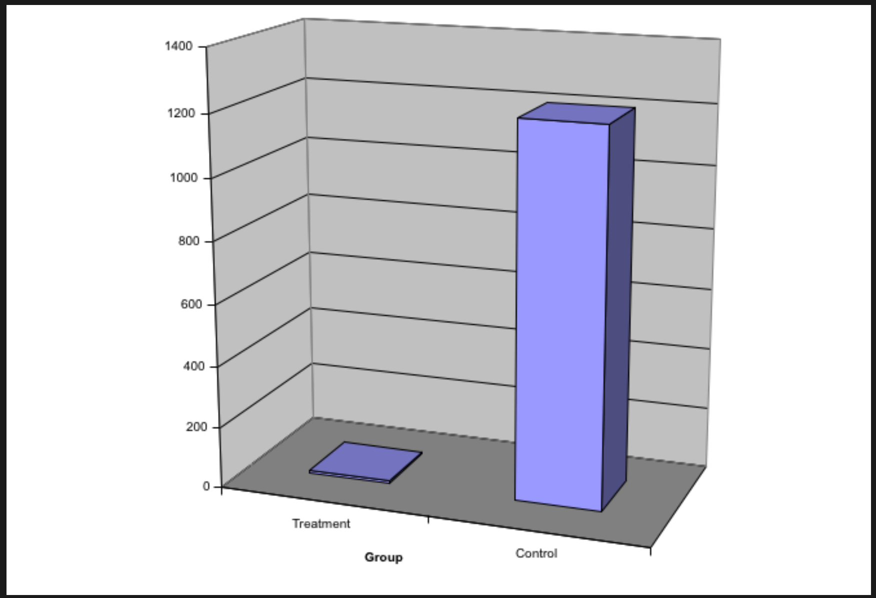
Consider logs



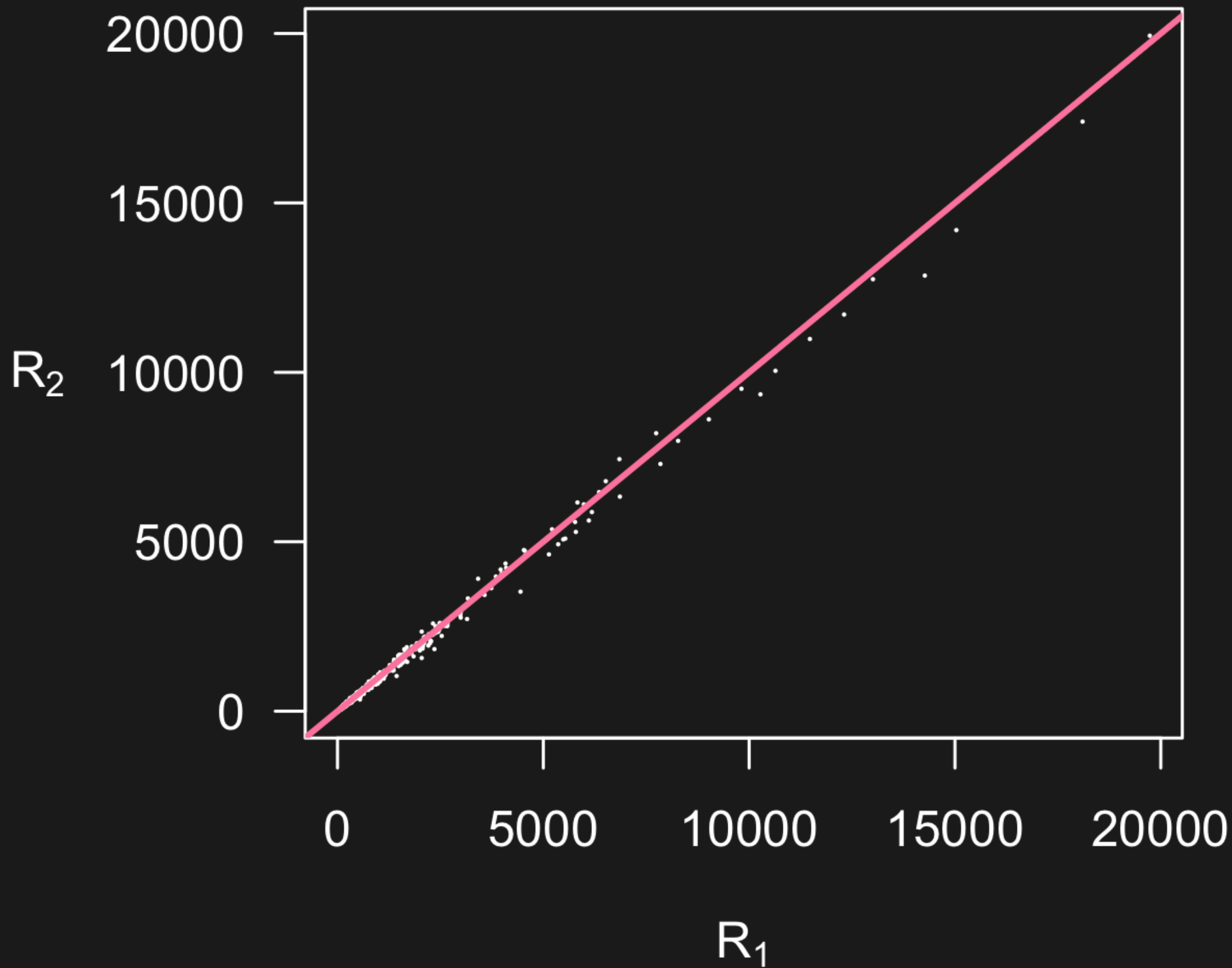


Consider logs

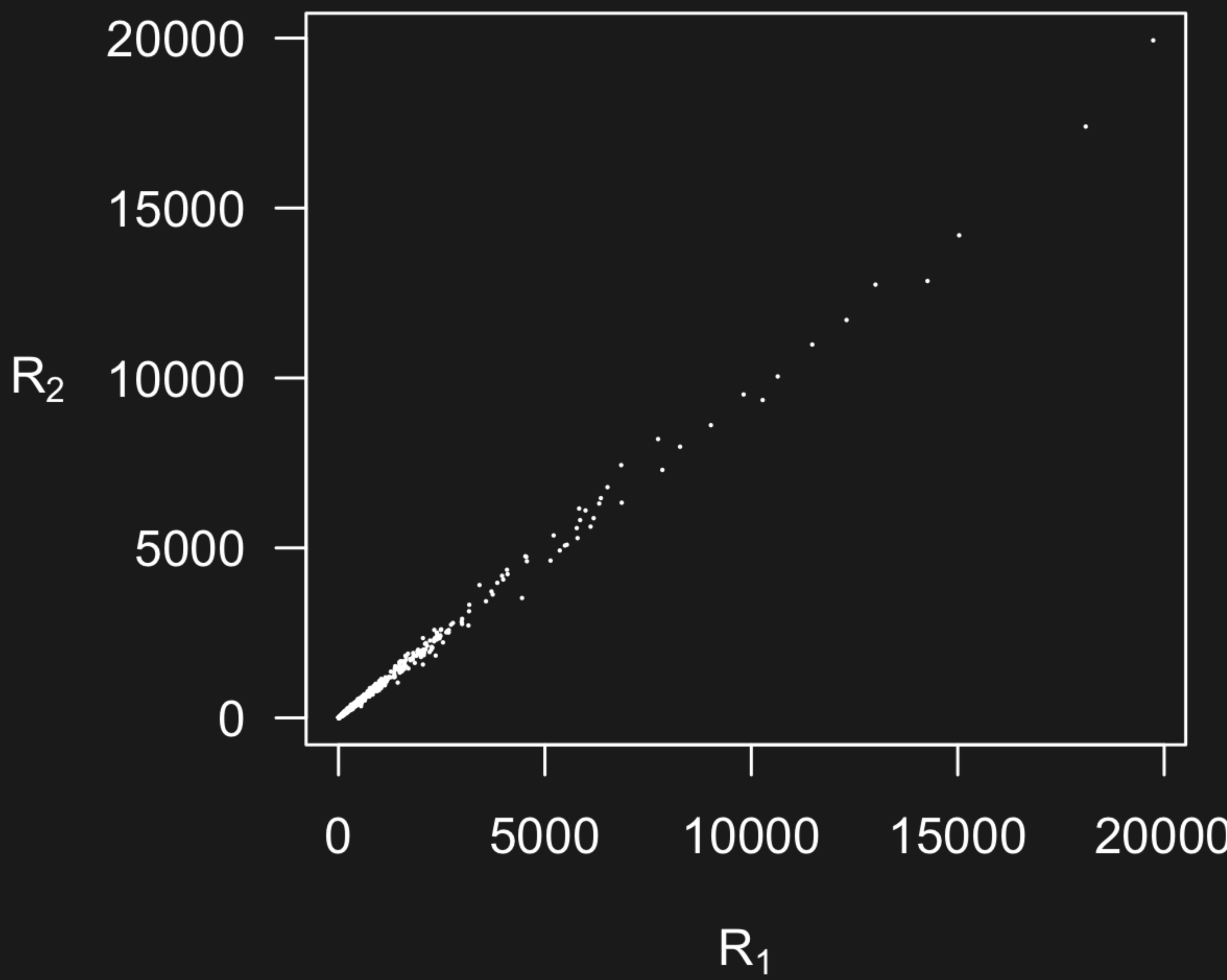




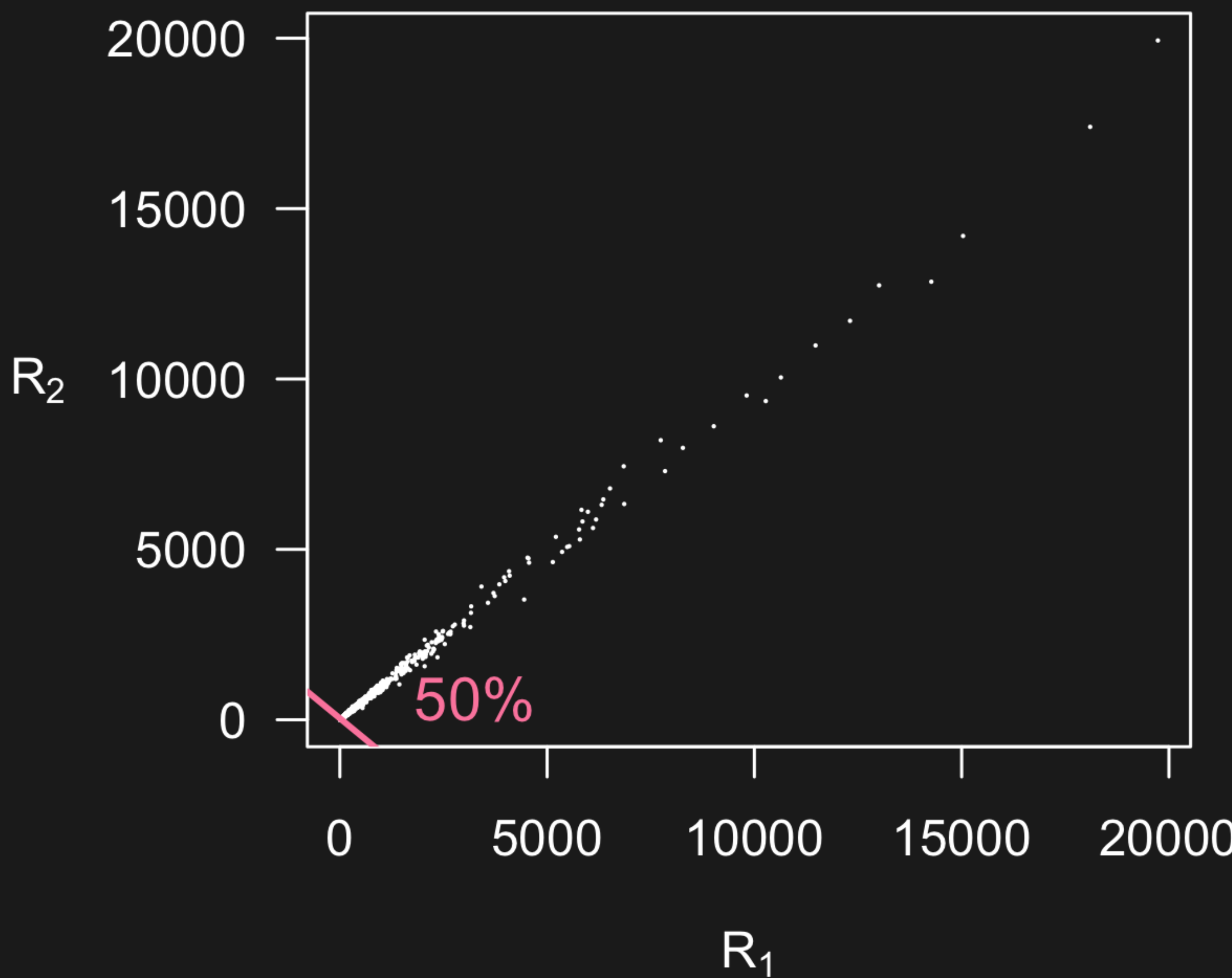
Consider logs



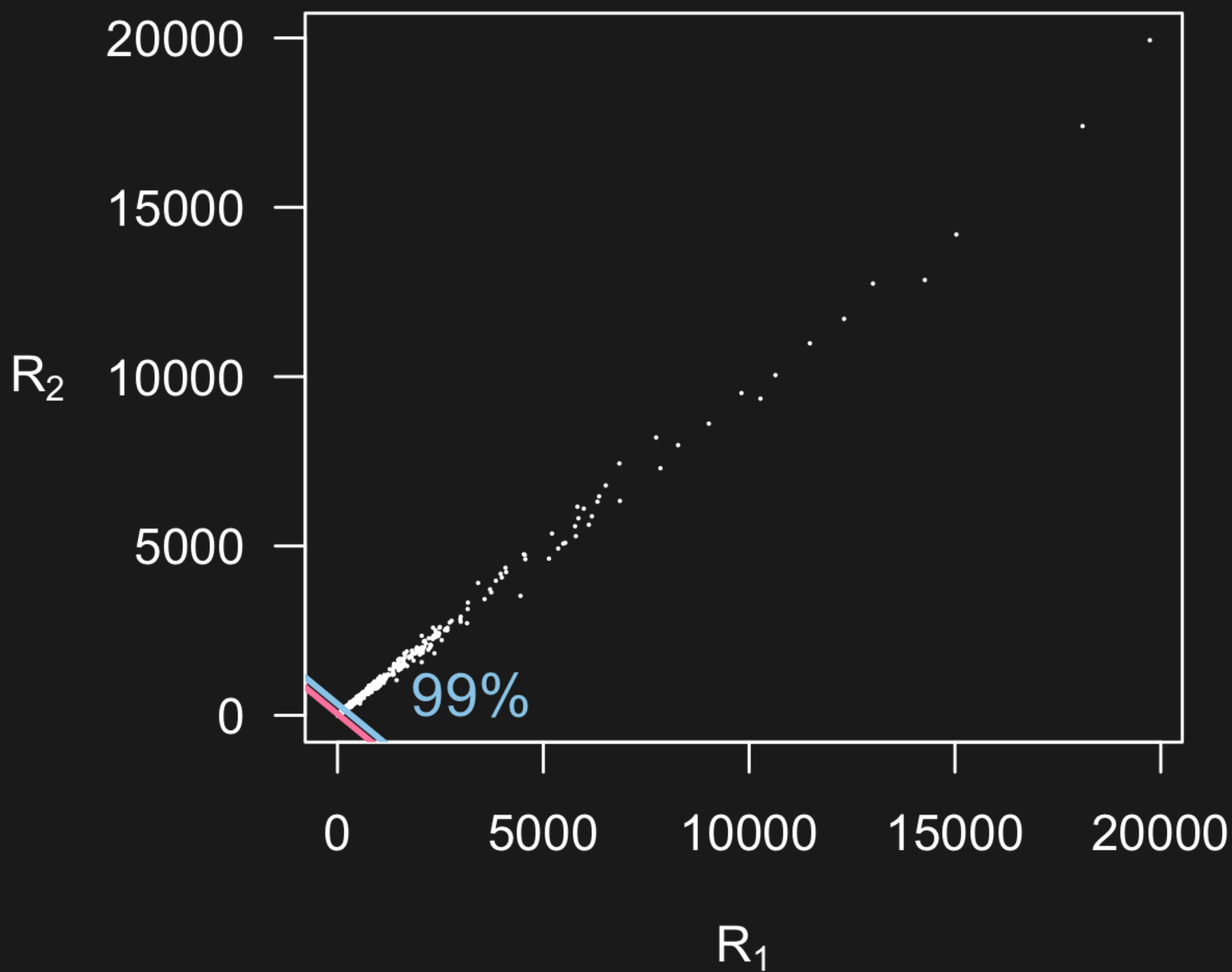
Consider logs



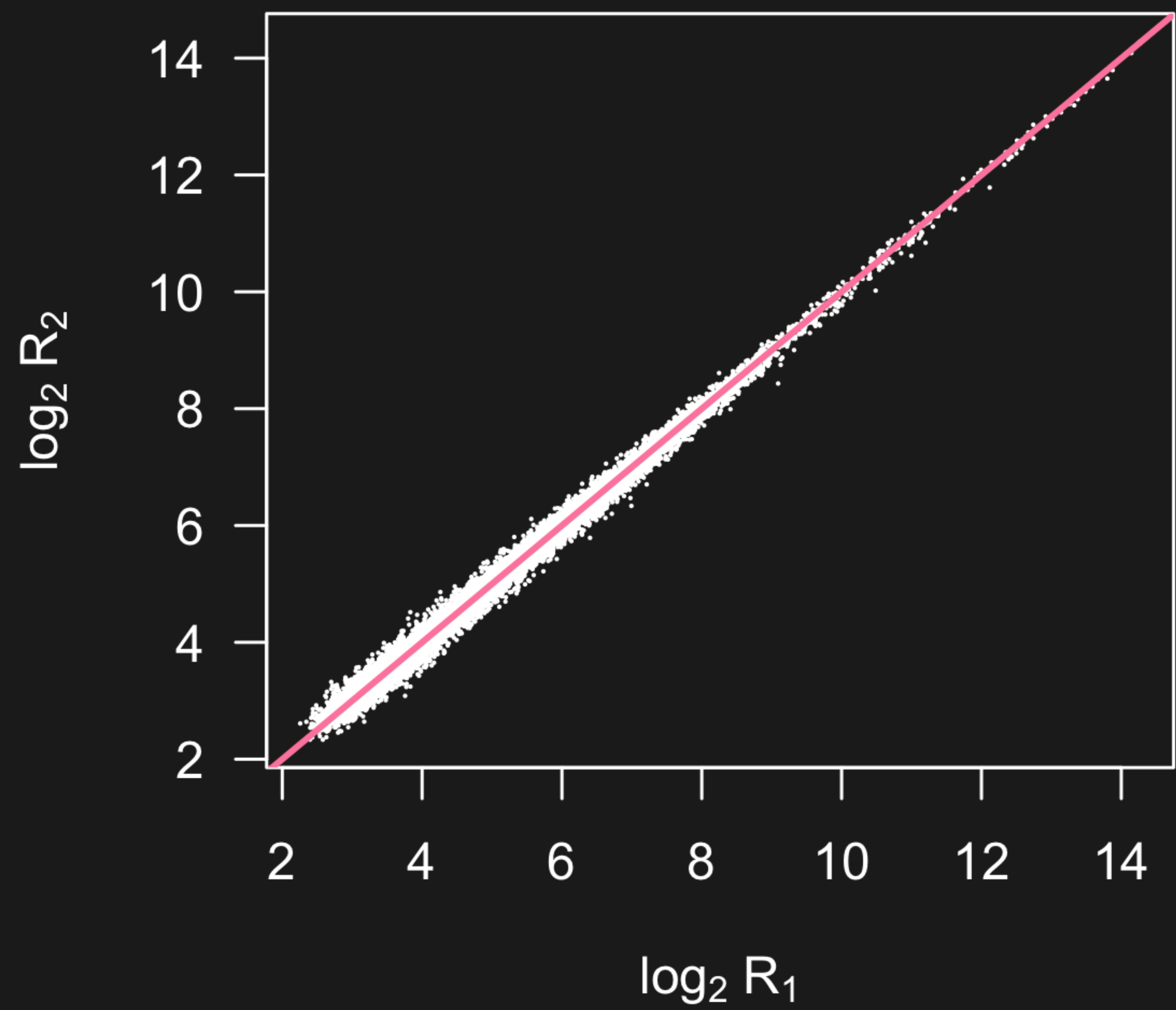
Consider logs



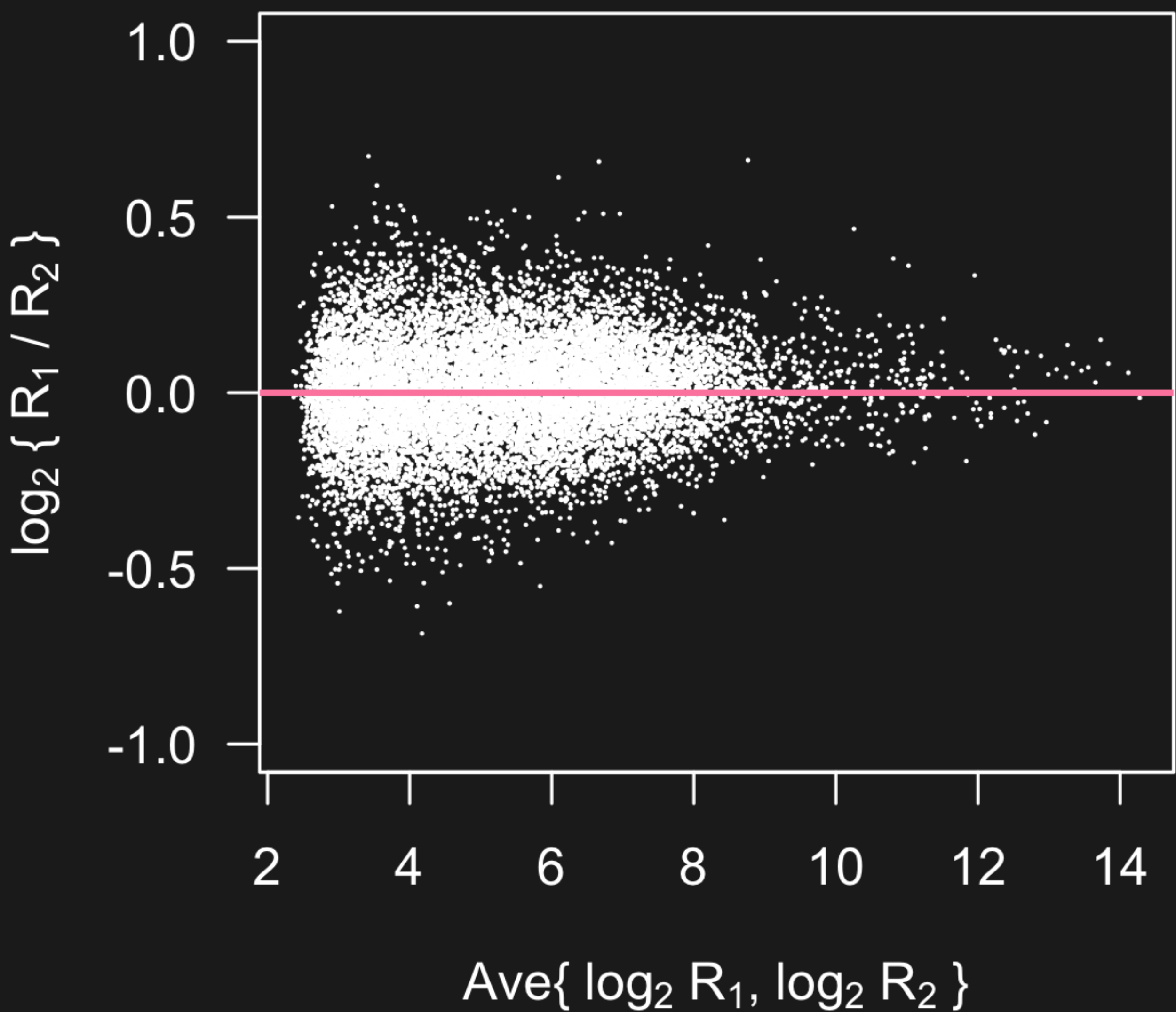
Consider logs



Consider logs

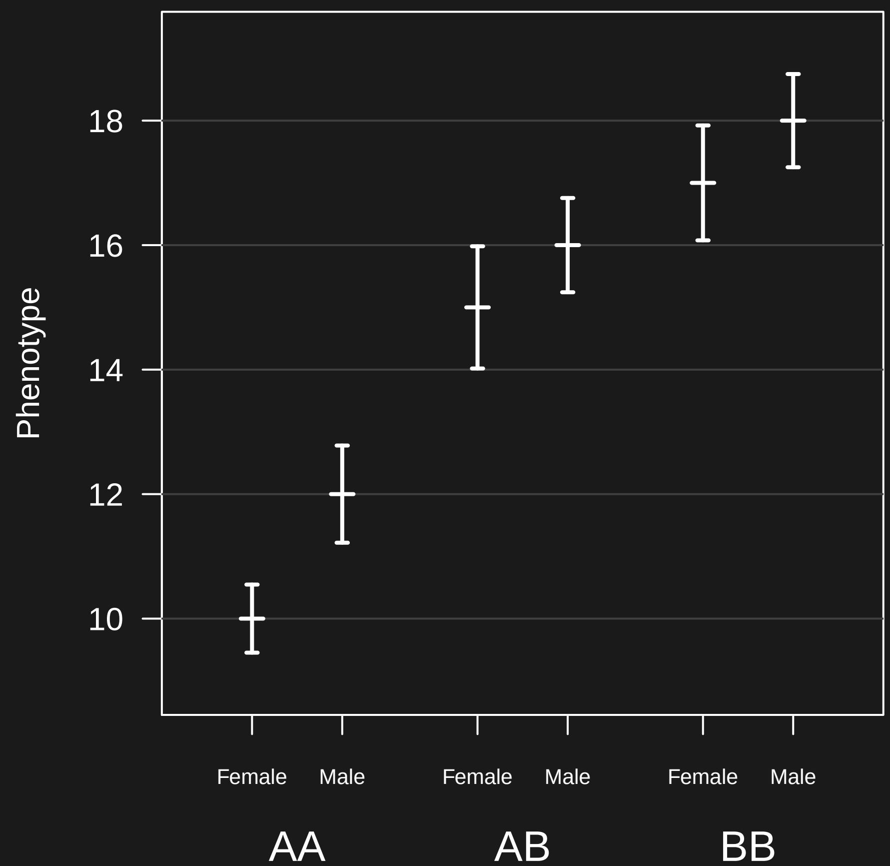


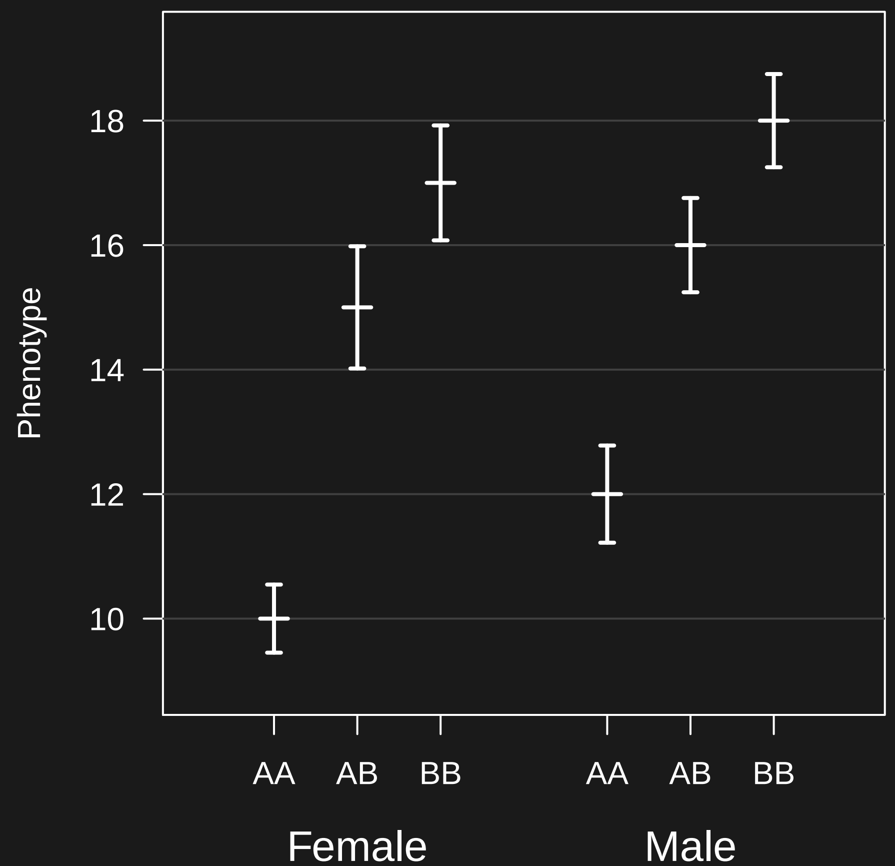
Take differences



Ease comparisons

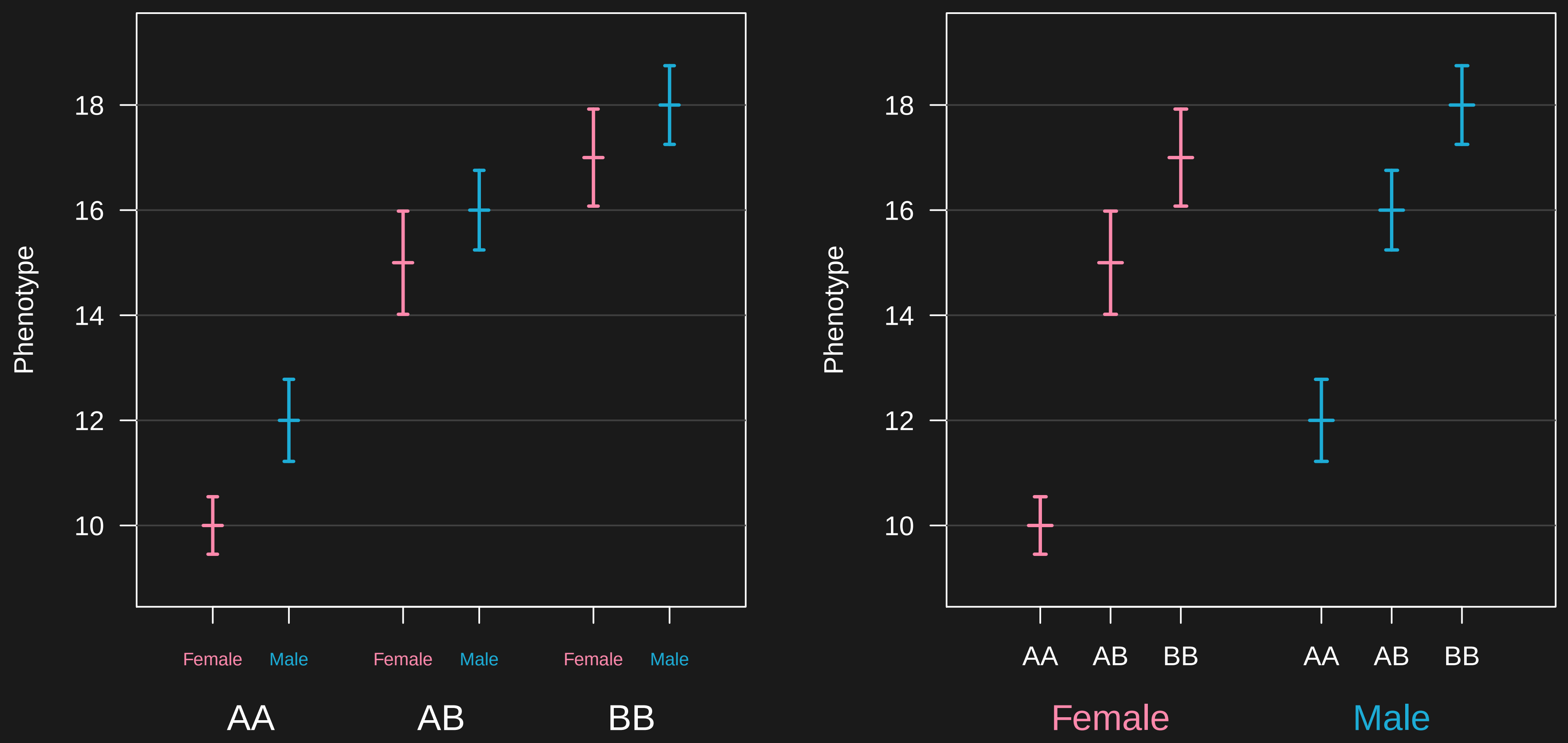
(things to be compared should be adjacent)



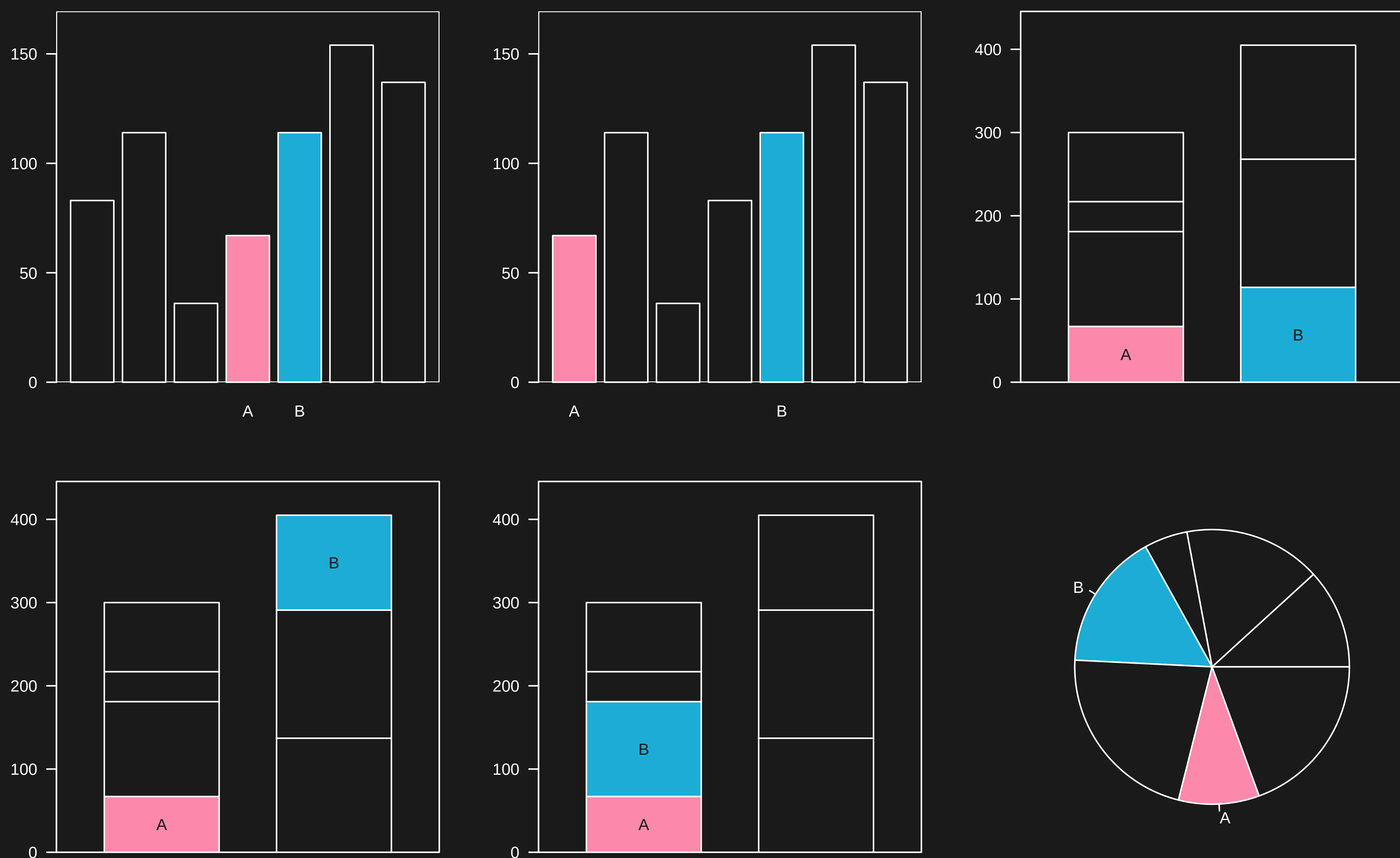


Ease comparisons

(add a bit of color)

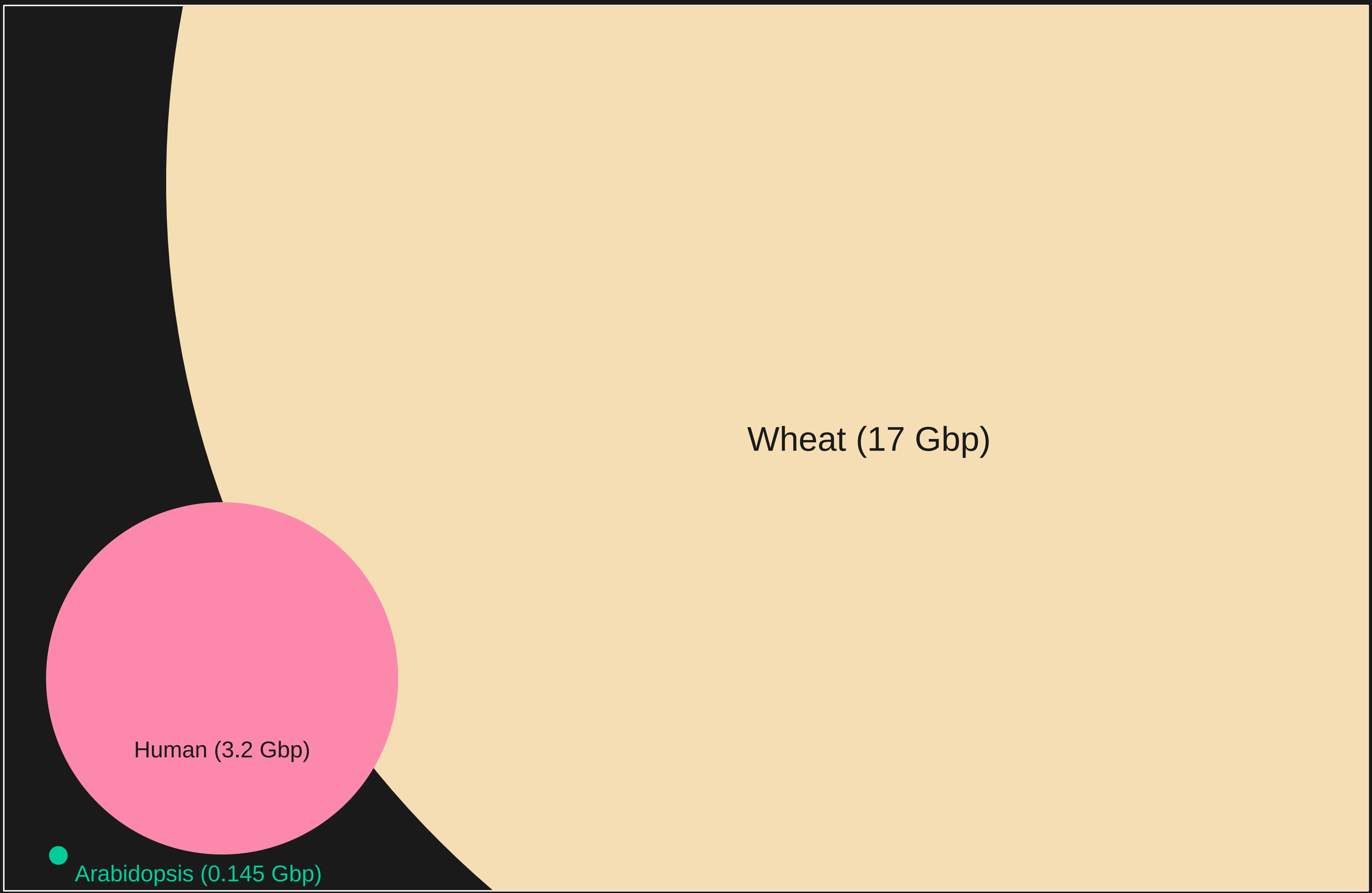


Which comparison is easiest?



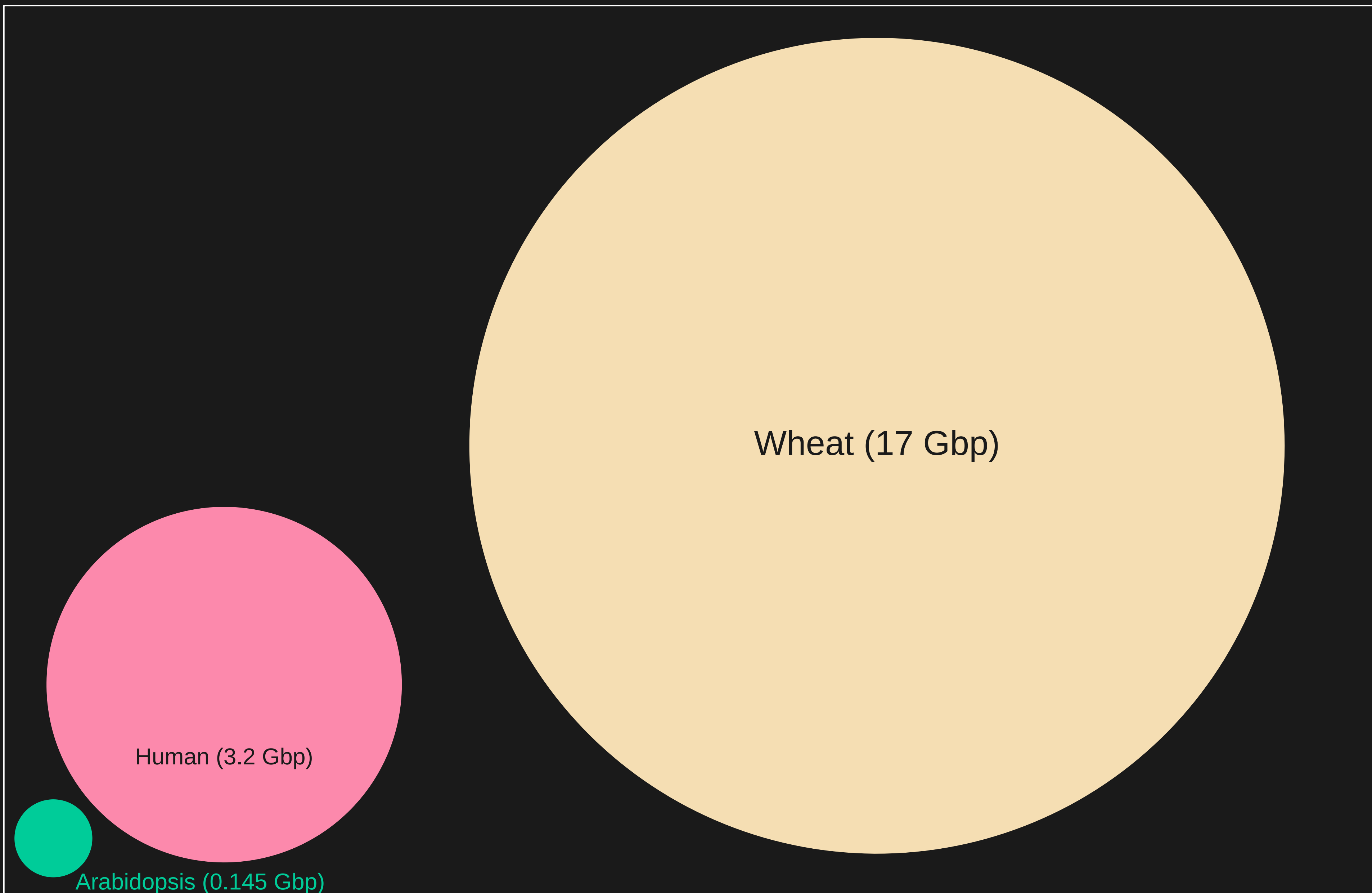
Don’t distort the quantities

(value $\propto$ radius)



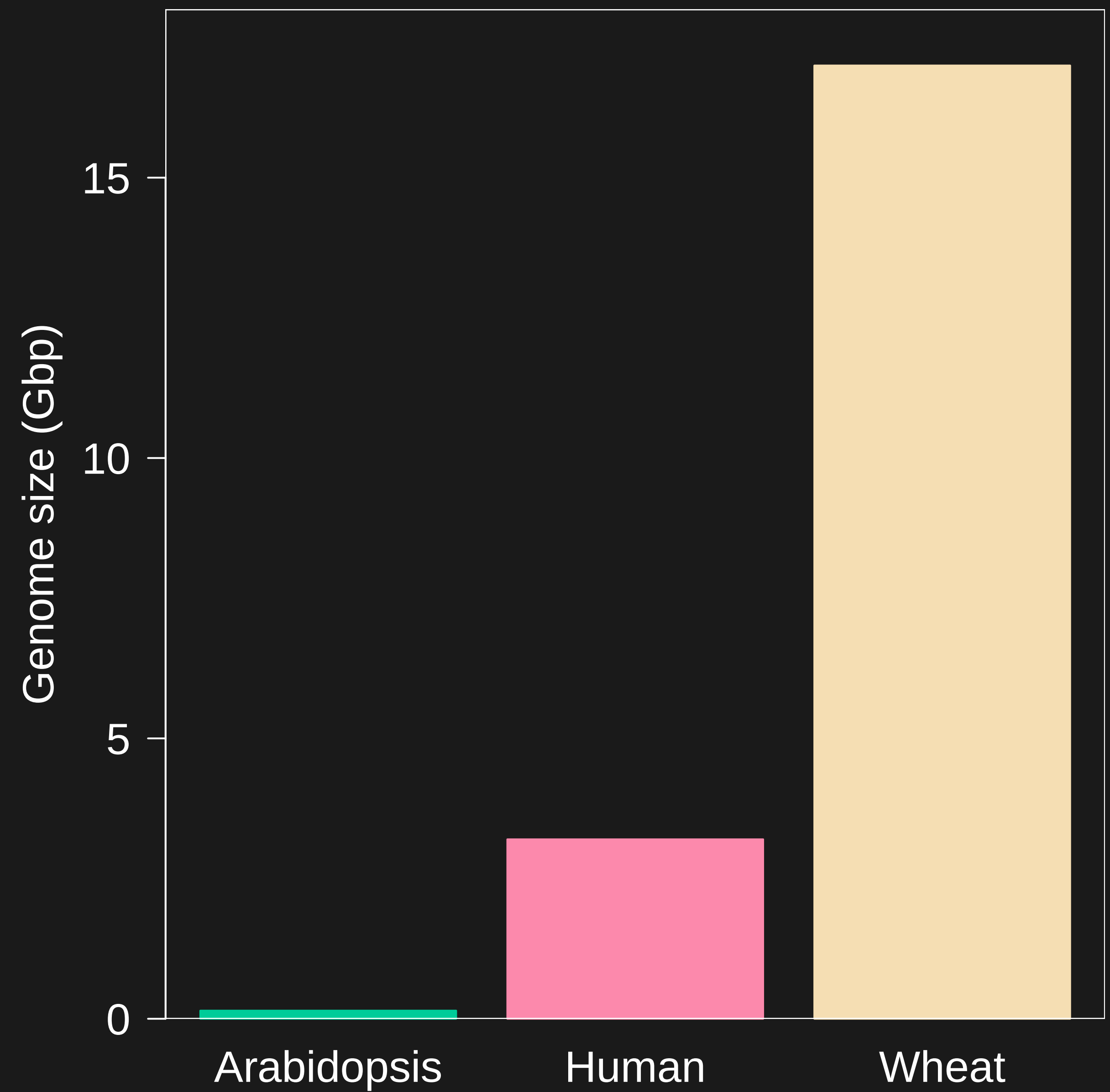
Don’t distort the quantities

(value $\propto$ area)



Don’t use areas at all

(value $\propto$ length)



Encoding data

Quantities

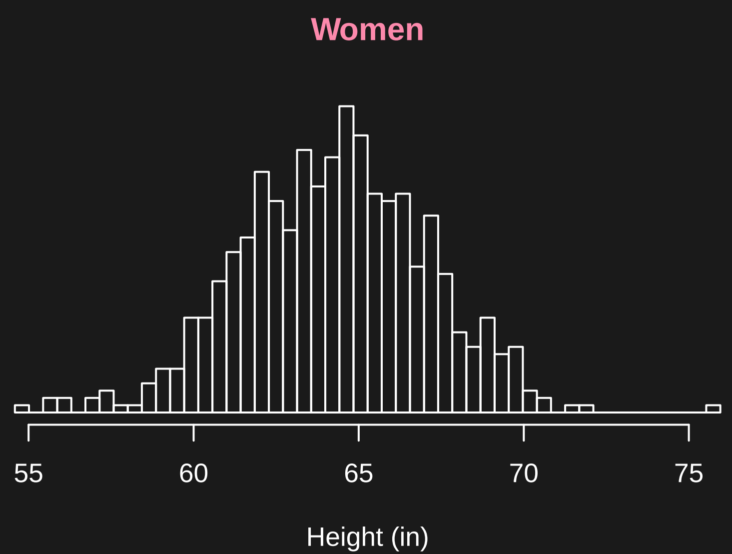
- Position
- Length
- Angle
- Area
- Luminance (light/dark)
- Chroma (amount of color)

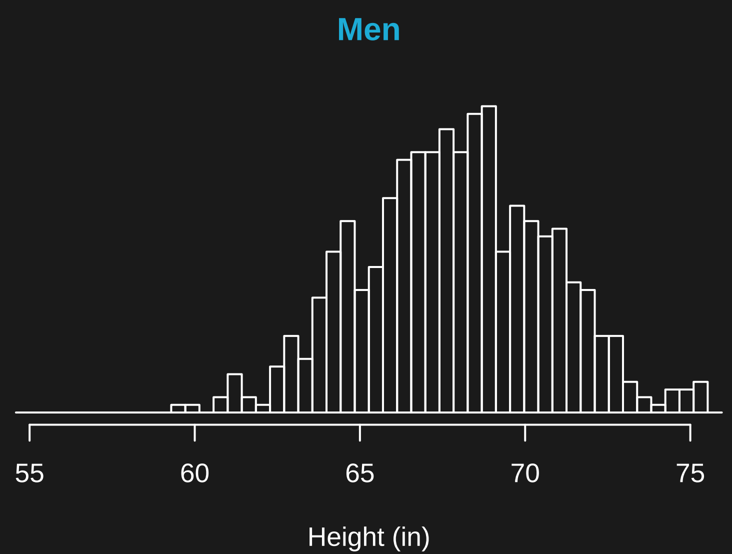
Categories

- Shape
- Hue (which color)
- Texture
- Width

Ease comparisons

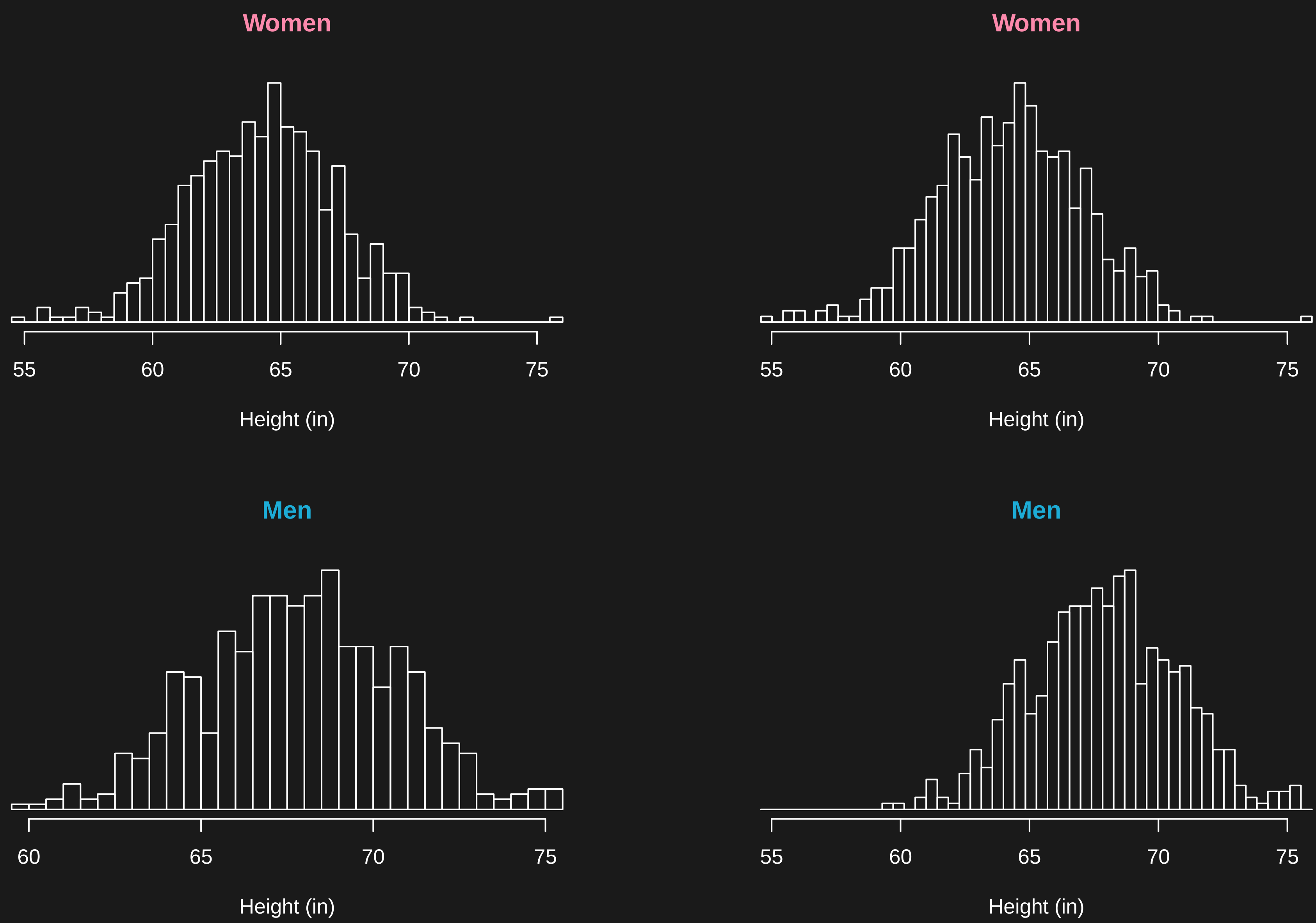
(align things vertically)



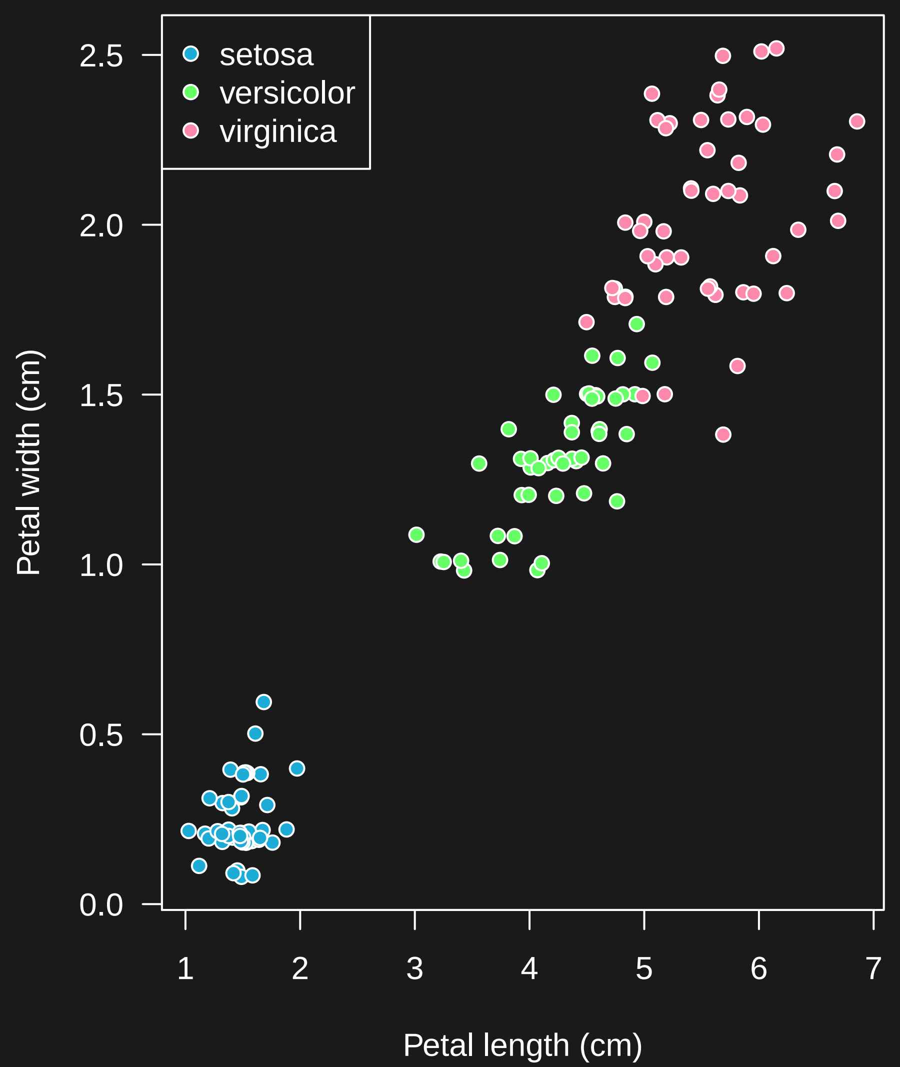


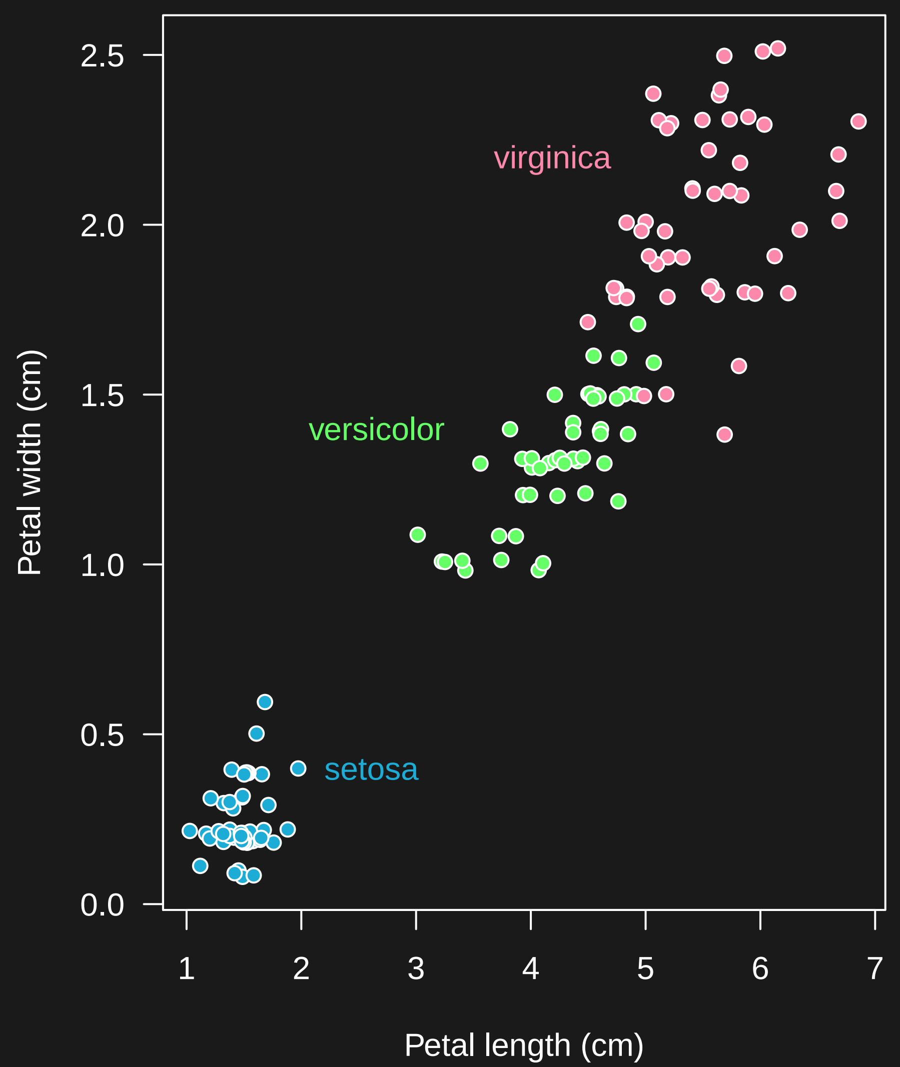
Ease comparisons

(use common axes)

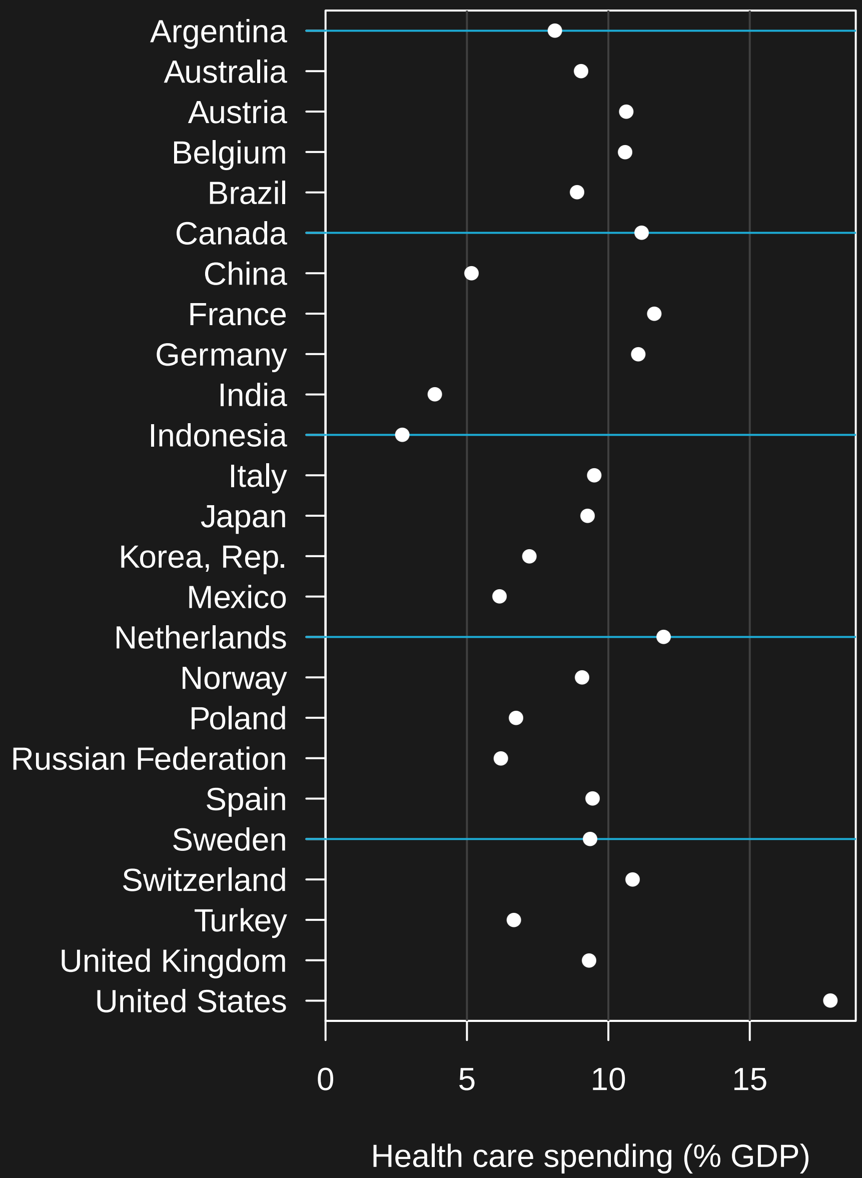


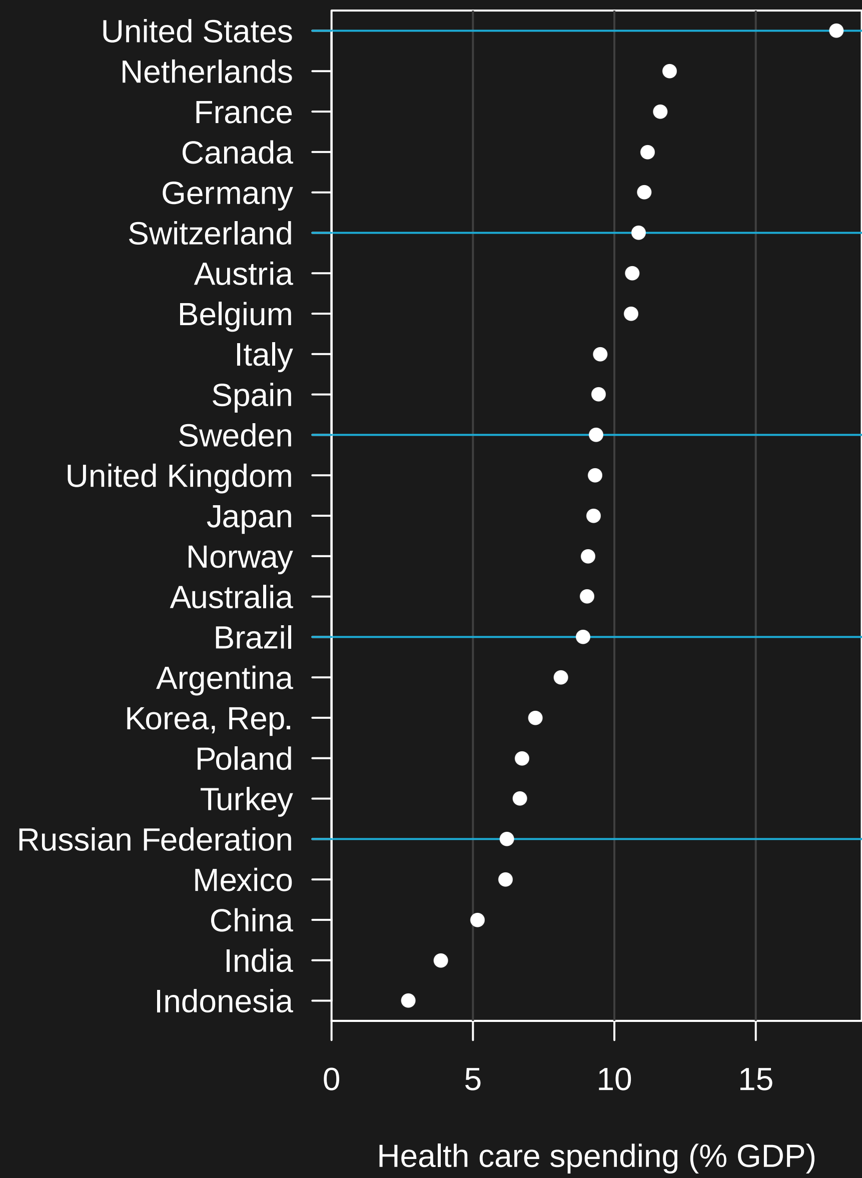
Use labels not legends



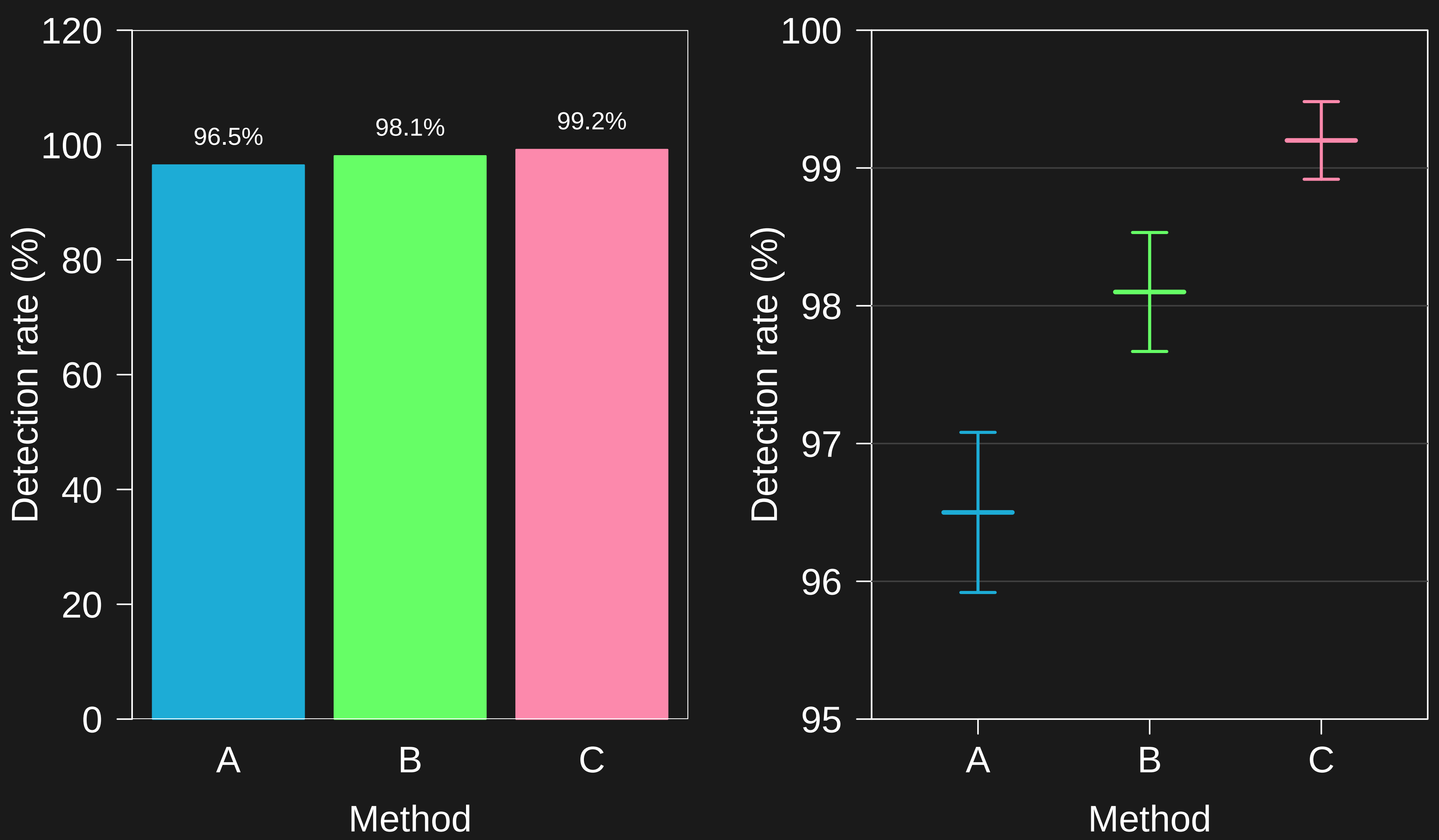


Don’t sort alphabetically





Must you include 0?



A bad table

	$b/c = 10.0$		$b/c = 10.0$		$b/c = 100.0$	
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

Fewer digits

	$b/c = 10.0$		$b/c = 10.0$		$b/c = 100.0$	
N	$r^\star$	G	$r^\star$	G	$r^\star$	G
3	2	0.20	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

Yuck!

	1990		2005		2010		p value
	n	Rate (95% CI)	n	Rate (95% CI)	n	Rate (95% CI)	
(Continued from previous page)							
Globally							
<75 years							
Incidence	6 353 868	159·22 (145·32–174·98)	9 288 048	167·45 (150·96–187·11)	10 469 624	168·75 (152·43–187·09)	0·208
Prevalence	13 234 062	324·26 (288·74–374·96)	20 187 246	358·58 (317·58–412·79)	23 052 804	366·93 (328·04–420·66)	0·086
MIR	..	0·359 (0·318–0·409)	..	0·293 (0·249–0·332)	..	0·254 (0·212–0·287)	<0·001
DALYs lost	63 991 864	1543·96 (1452·03–1728·25)	74 855 520	1326·17 (1172·08–1388·74)	73 293 552	1163·448 (1011·43–1232·19)	<0·001
Mortality	2 301 435	57·38 (54·12–64·27)	2 734 251	49·16 (43·60–51·55)	2 668 499	42·89 (37·65–45·81)	<0·001
≥75 years							
Incidence	3 725 067	3173·50 (2932·14–3422·23)	5 446 077	3082·97 (2819·52–3372·55)	6 424 911	3113·00 (2850·95–3403·57)	0·361
Prevalence	4 681 276	3974·37 (3609·66–4441·23)	8 308 337	4700·18 (4239·37–5256·84)	9 972 153	4835·38 (4382·63–5433·92)	0·005
MIR	..	0·634 (0·575–0·709)	..	0·543 (0·476–0·607)	..	0·500 (0·439–0·560)	<0·001
DALYs	22 018 520	18665·35 (17 464·55–20 408·51)	27 096 178	15 300·36 (13 987·78–16 317·62)	28 938 754	14 053·63 (12 761·98–15 088·12)	<0·001
Mortality	2 359 013	2033·21 (1888·78–2233·65)	2 950 719	1678·65 (1528·60–1807·22)	3 205 682	1545·29 (1412·76–1685·12)	<0·001
All ages							
Incidence	10 078 935	250·55 (229·70–273·25)	14 734 124	255·79 (232·10–283·88)	16 894 536	257·96 (234·40–284·11)	0·335
Prevalence	17 915 338	434·86 (389·45–496·84)	28 495 582	490·13 (436·60–557·52)	33 024 958	502·32 (451·26–572·18)	0·047
MIR	..	0·461 (0·415–0·518)	..	0·386 (0·336–0·432)	..	0·348 (0·299–0·390)	<0·001
DALYs lost	86 010 384	2062·74 (1949·53–2280·29)	101 951 696	1749·59 (1568·67–1830·82)	102 232 304	1554·02 (1373·94–1642·26)	<0·001
Mortality	4 660 449	117·25 (111·51–129·68)	5 684 970	98·53 (89·02–103·86)	5 874 182	88·41 (79·84–94·41)	<0·001

*p value for the difference in age-adjusted rates between 1990 and 2010 only.

Table 1: **Age-adjusted annual incidence and mortality rates (per 100 000 person-years), disability-adjusted life-years (DALYs) lost, prevalence (per 100 000 people), and mortality-to-incidence ratio (MIR) by age groups in high-income and low-income and middle-income countries, and globally in 1990, 2005, and 2010**

Feigen et al., Lancet 383:245-255, 2014, Table 1

Yuck!

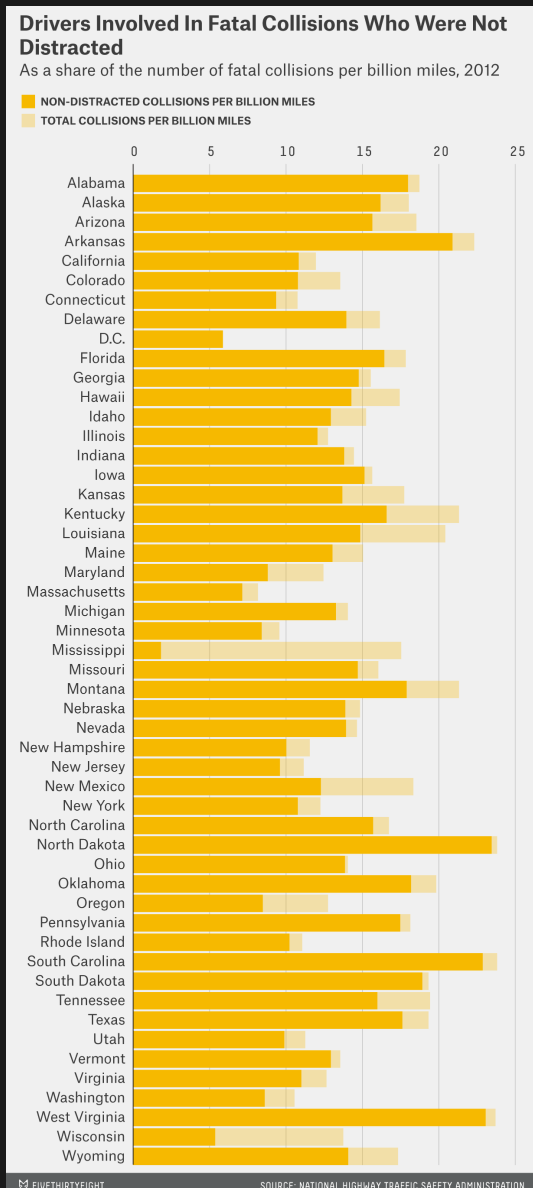
	1990	
	n	Rate (95% CI)
(Continued from previous page)		
Globally		
<75 years		
Incidence	6 353 868	159·22 (145·32–174·98)
Prevalence	13 234 062	324·26 (288·74–374·96)
MIR	..	0·359 (0·318–0·409)
DALYs lost	63 991 864	1543·96 (1452·03–1728·25)
Mortality	2 301 435	57·38 (54·12–64·27)

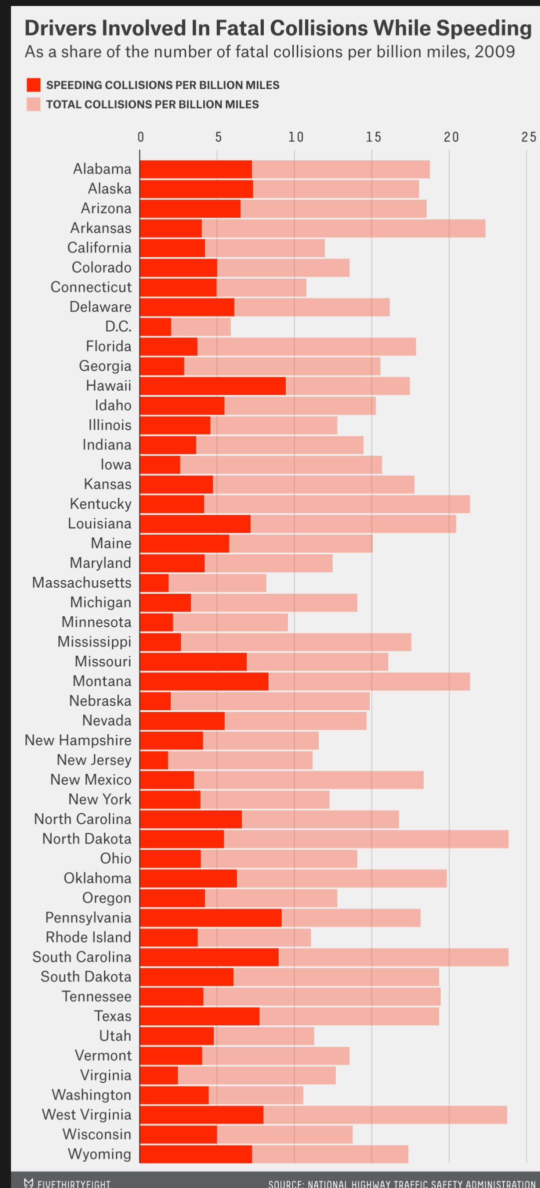
Feigen et al., Lancet 383:245-255, 2014, Table 1

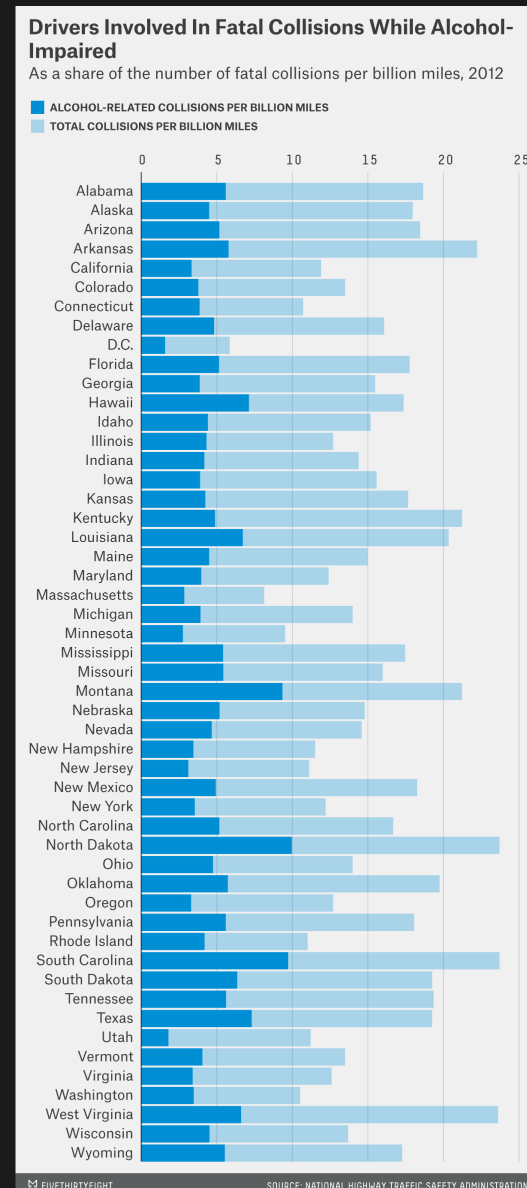
What was wrong with that?

- Way too many digits.
- Numbers aren't aligned.
- Numbers to be compared aren't anywhere near each other.
- The interesting comparisons are horizontal rather than vertical.
- It would be much better as a multi-panel figure.

One last example

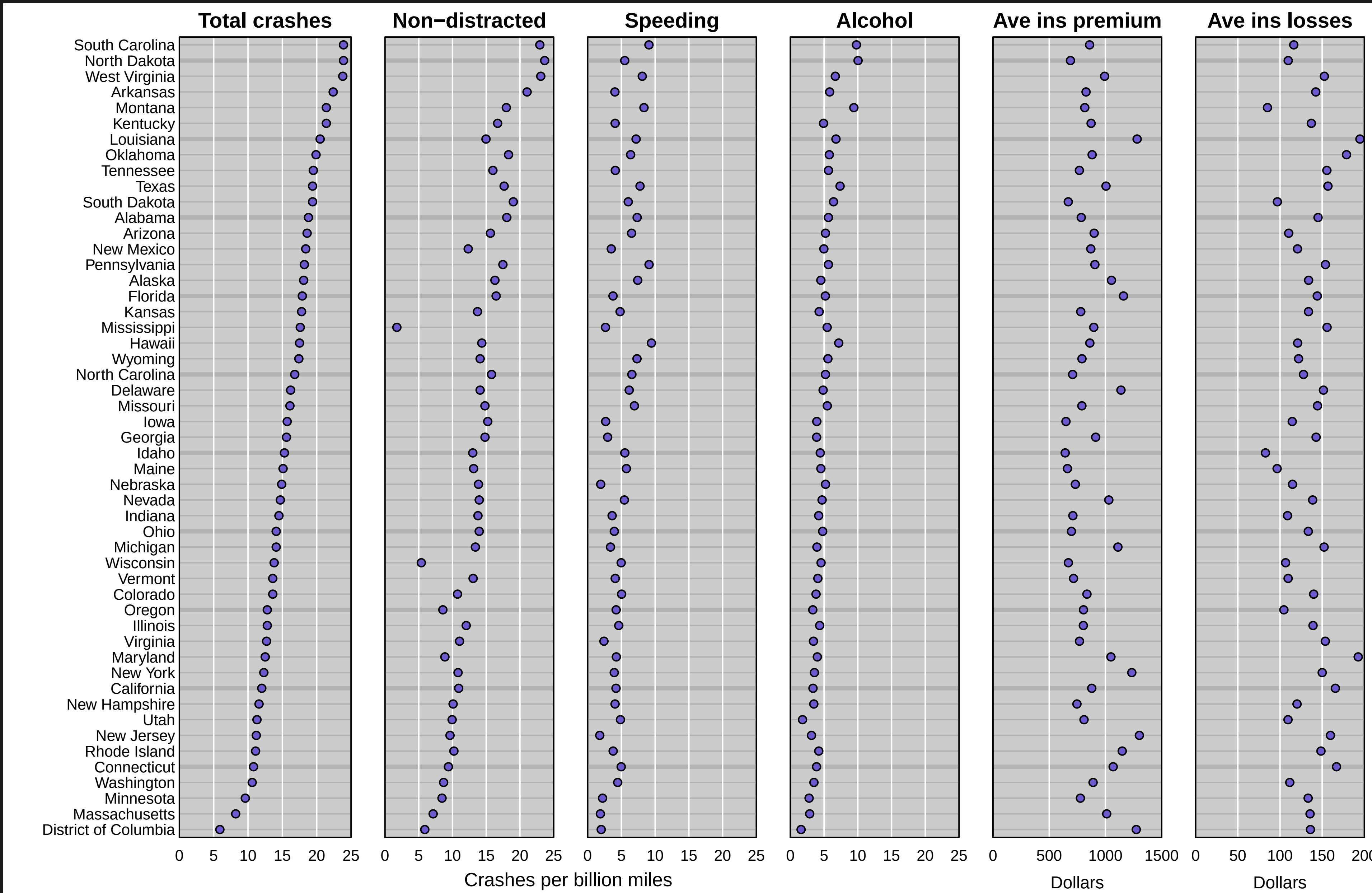




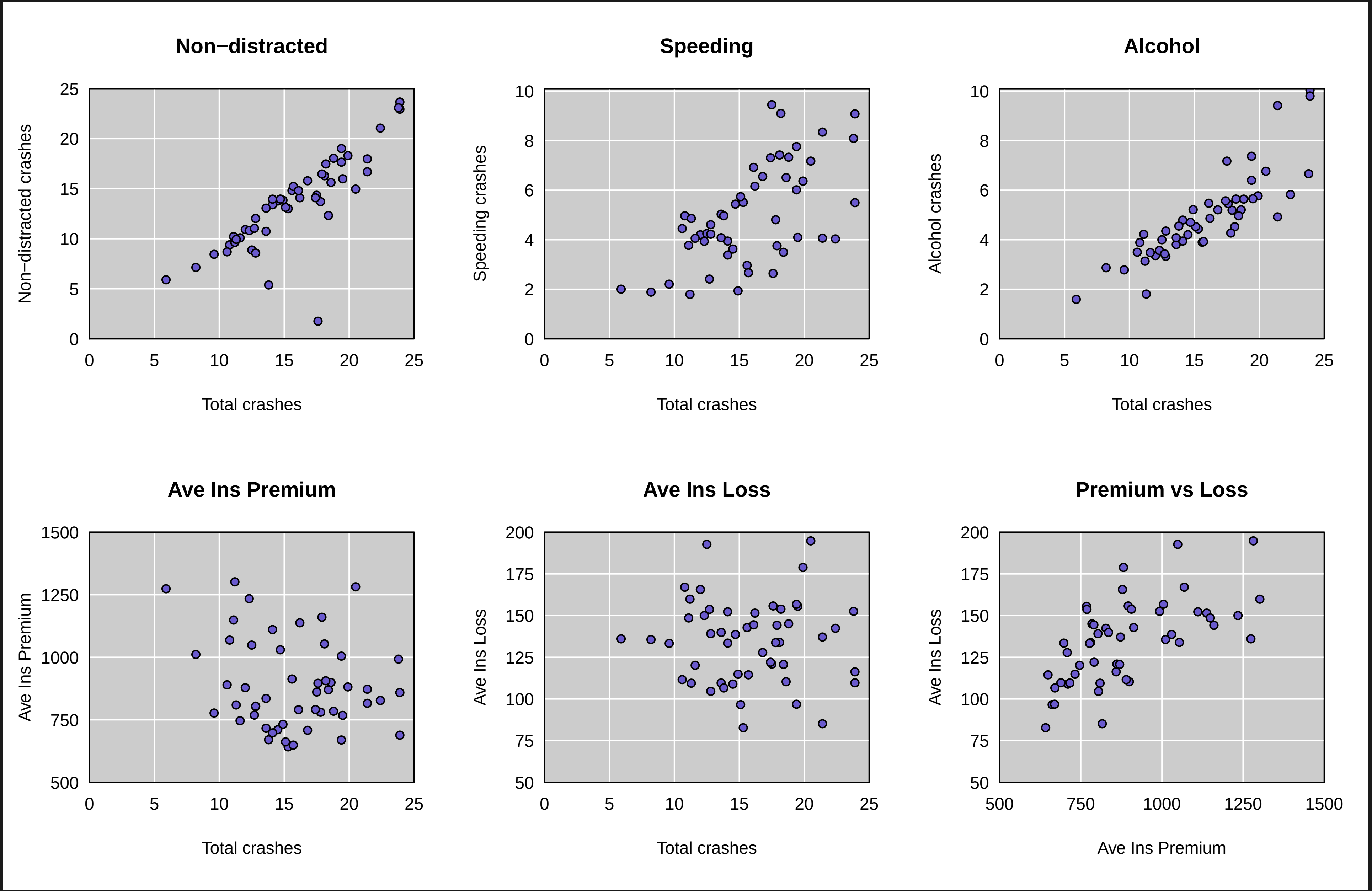


fivethirtyeight.com/datalab/which-state-has-the-worst-drivers

An alternative



Scatterplots



Summary I

- Show the data
- Avoid chart junk
- Consider taking logs and/or differences
- Put the things to be compared next to each other
- Use color to set things apart, but consider color blind folks
- Use position rather than angle or area to represent quantities

Summary II

- Align things vertically to ease comparisons
- Use common axis limits to ease comparisons
- Use labels rather than legends
- Sort on meaningful variables (not alphabetically)
- Must 0 be included in the axis limits?
- Use scatterplots to explore relationships

Inspirations

- Hadley Wickham (slides at `http://courses.had.co.nz`)
- Naomi Robbins (*Creating more effective graphs*)
- Howard Wainer
- Andrew Gelman
- Dan Carr
- Edward Tufte

Further reading

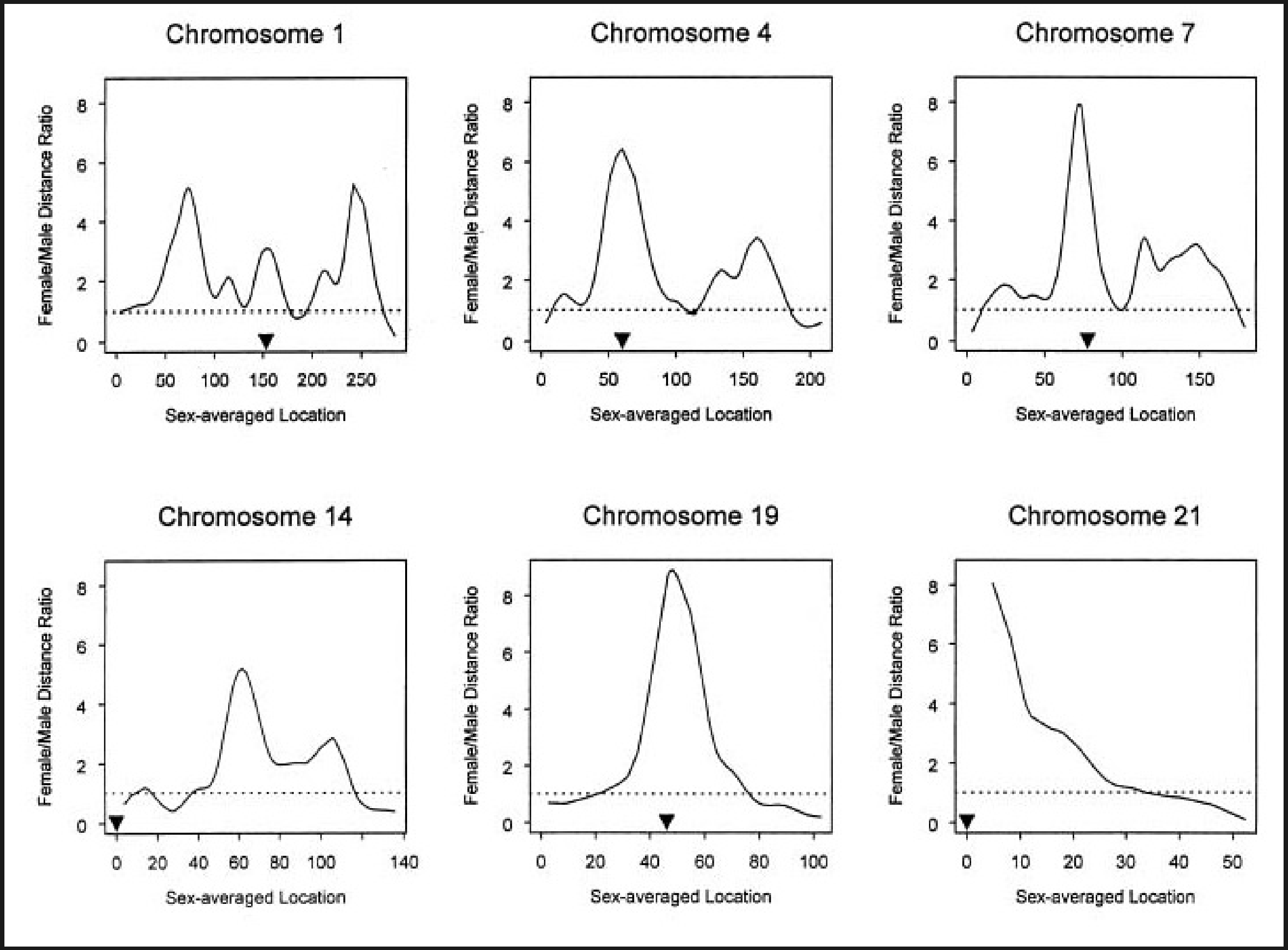
- ER Tufte (1983) The visual display of quantitative information. Graphics Press.
- ER Tufte (1990) Envisioning information. Graphics Press.
- ER Tufte (1997) Visual explanations. Graphics Press.
- A Gelman, C Pasarica, R Dodhia (2002) Let's practice what we preach: Turning tables into graphs. The American Statistician 56:121-130
- NB Robbins (2004) Creating more effective graphs. Wiley
- Nature Methods columns: `http://bang.clearscience.info/?p=546`
- These slides: `bit.ly/CDHA2016`

The top ten worst graphs

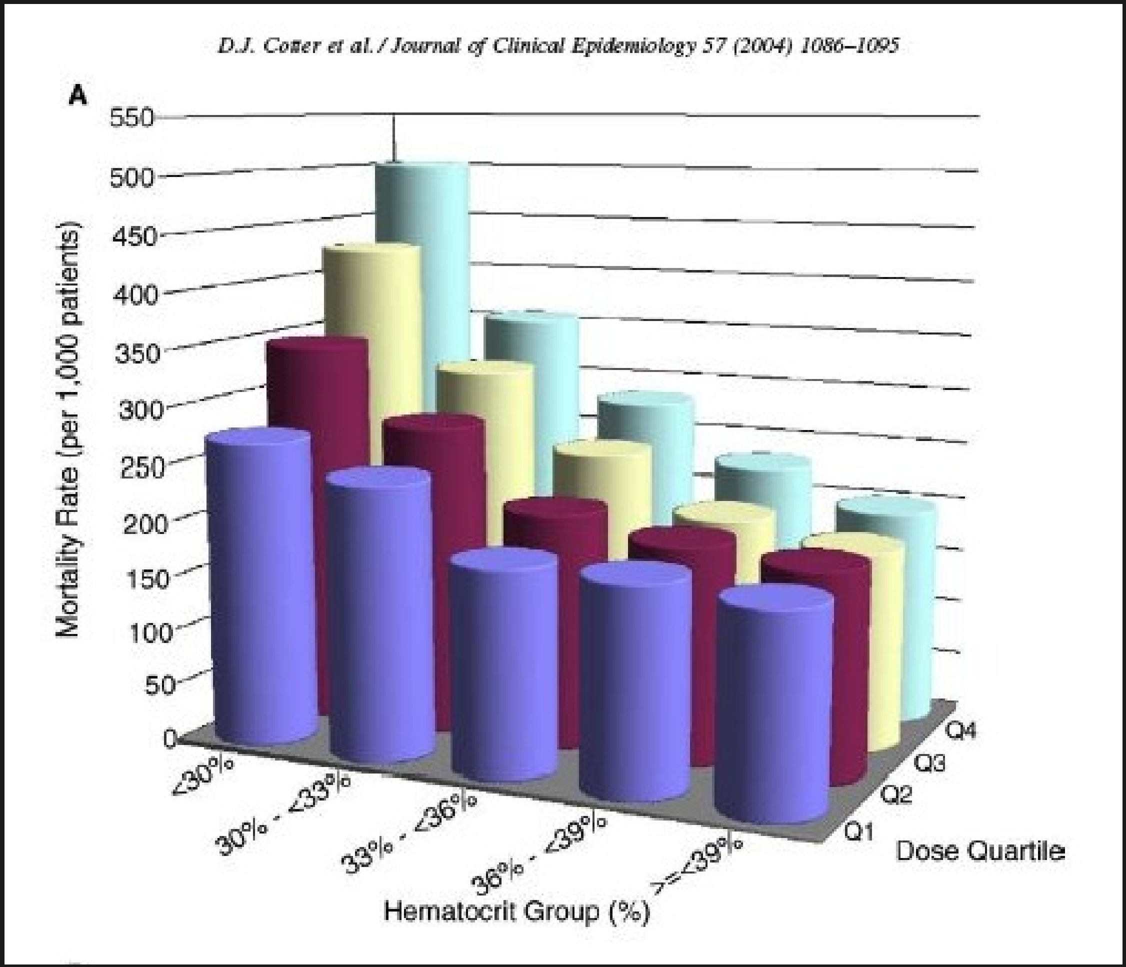
With apologizes to the authors, we provide the following list of the top ten worst graphs in the scientific literature.

As these examples indicate, good scientists can make mistakes.

`bit.ly/TopTenWorstGraphs`

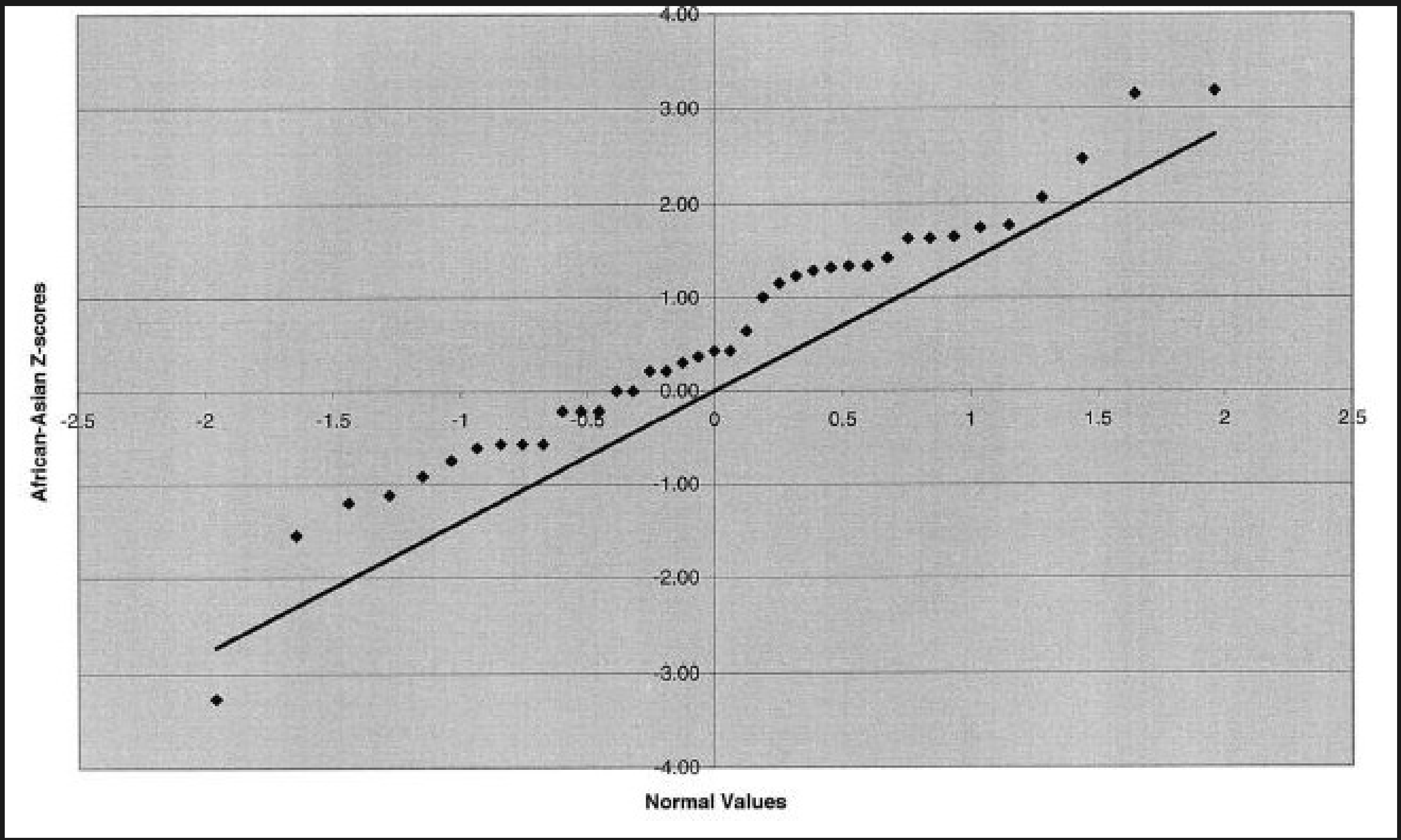


Broman et al., Am J Hum Genet 63:861-869, 1998, Fig. 1

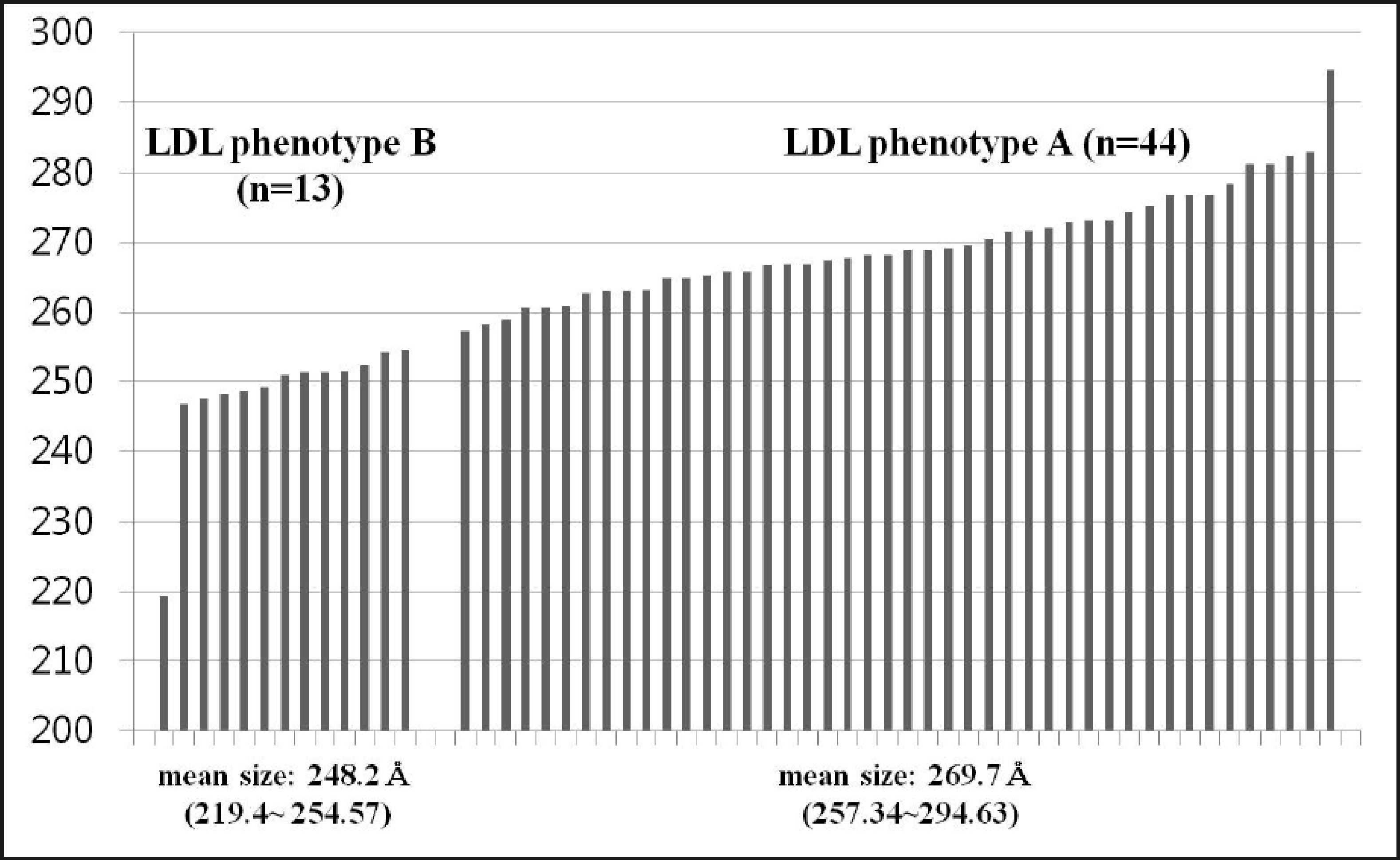


Cotter et al., J Clin Epidemiol 57:1086-1095, 2004, Fig 2

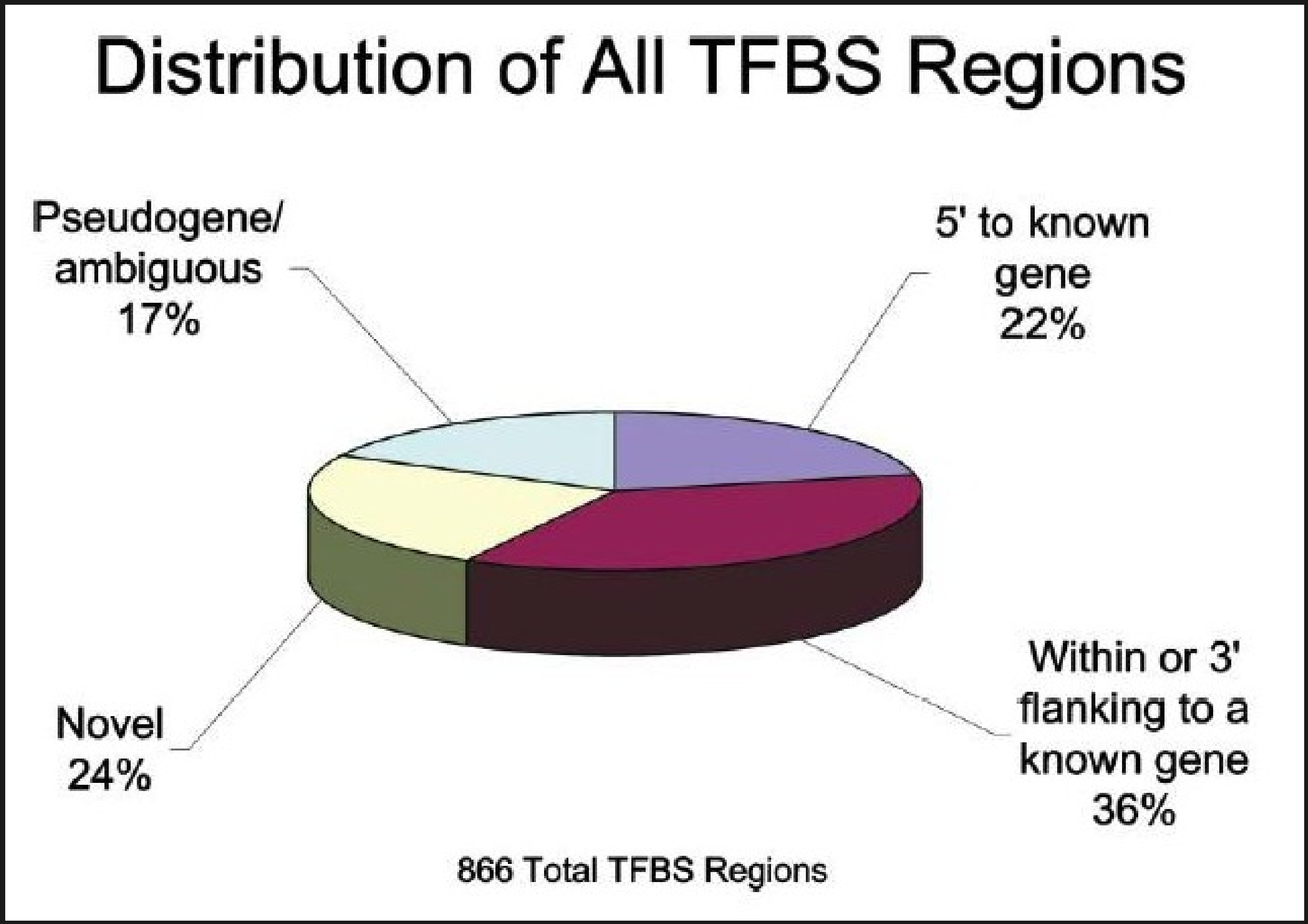
8



Jorgenson et al., Am J Hum Genet 76:276-290, 2005, Fig 2

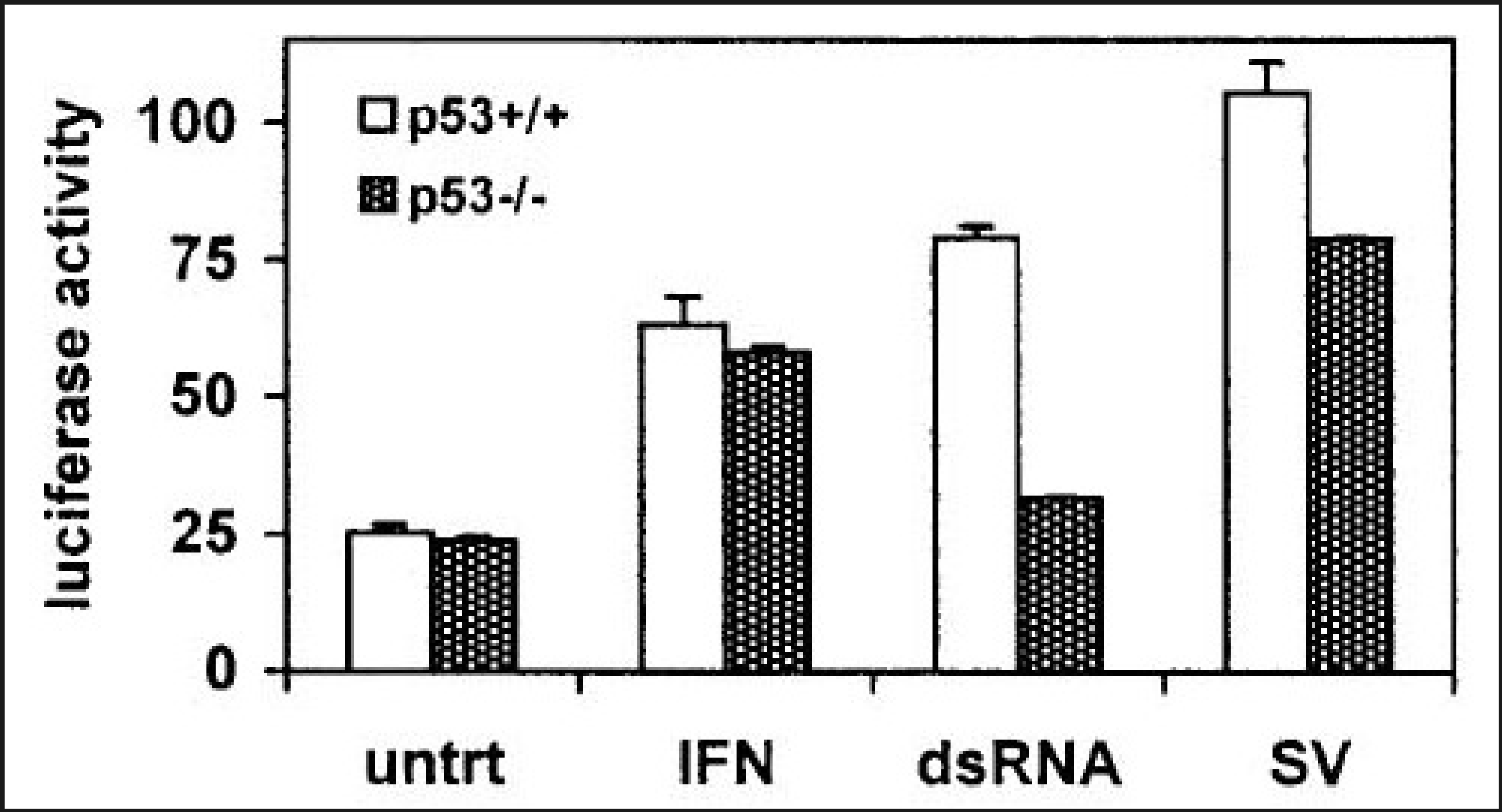


Kim et al., Nutr Res Prac 6:120-125, 2012 Fig 1



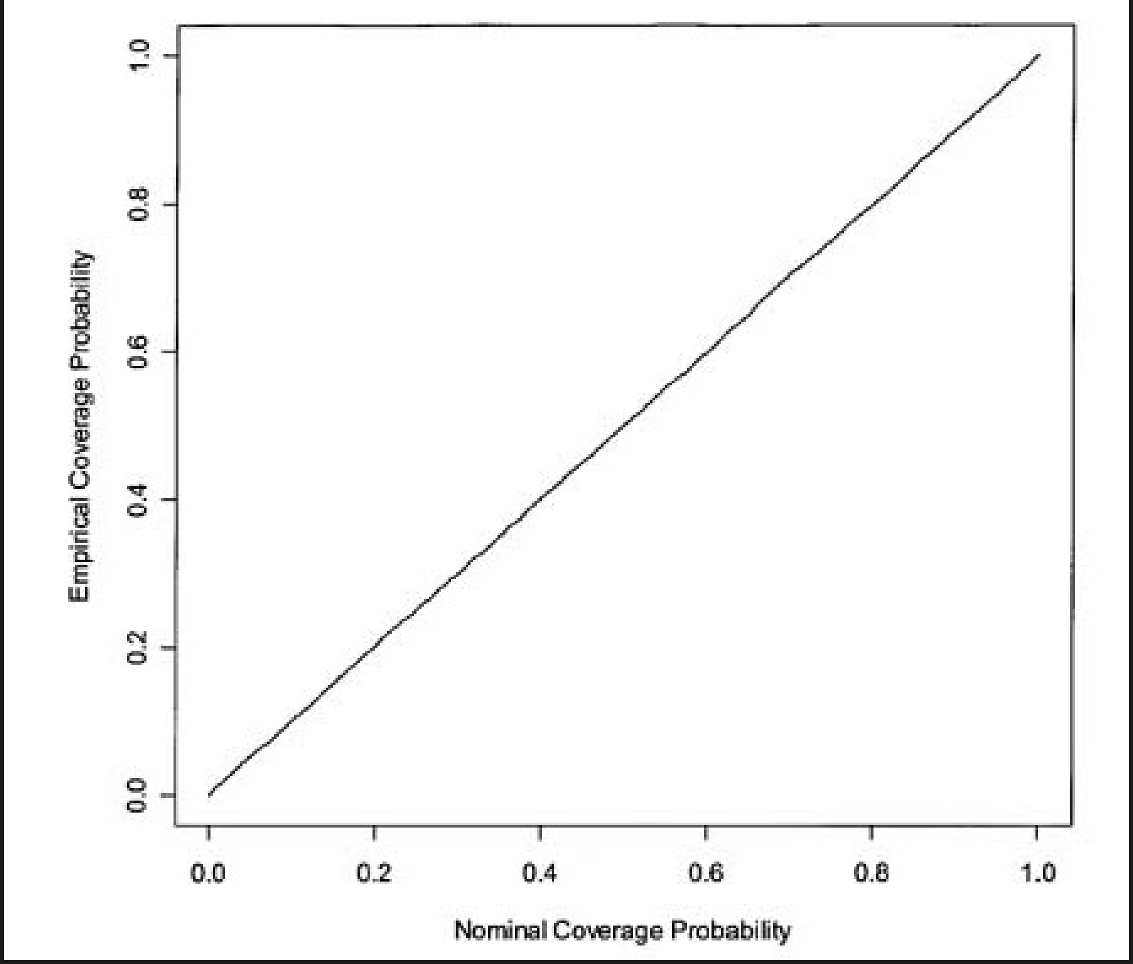
Cawley et al., Cell 116:499-509, 2004, Fig 1

5

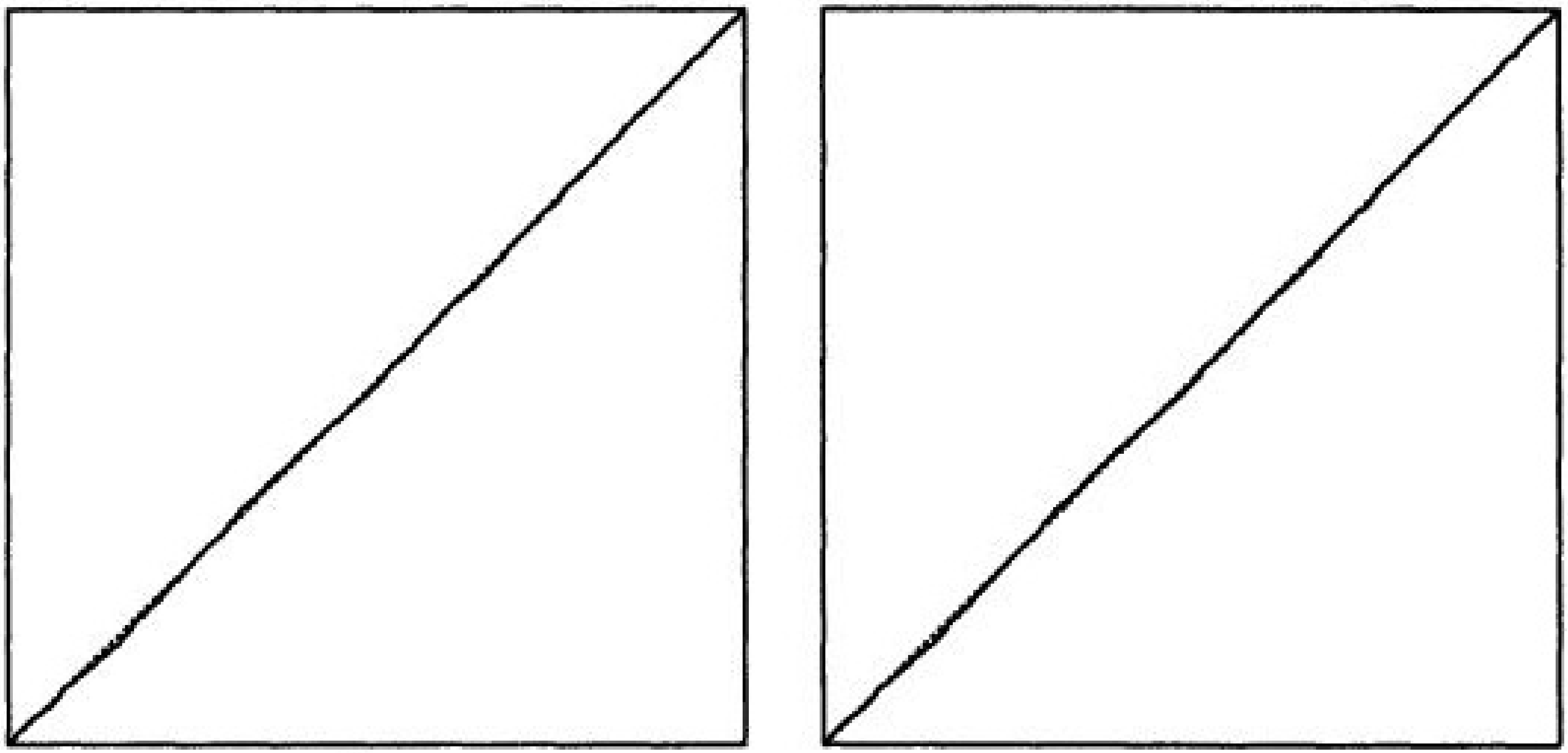


Hummer et al., J Virol 75:7774-7777, 2001, Fig 4

4

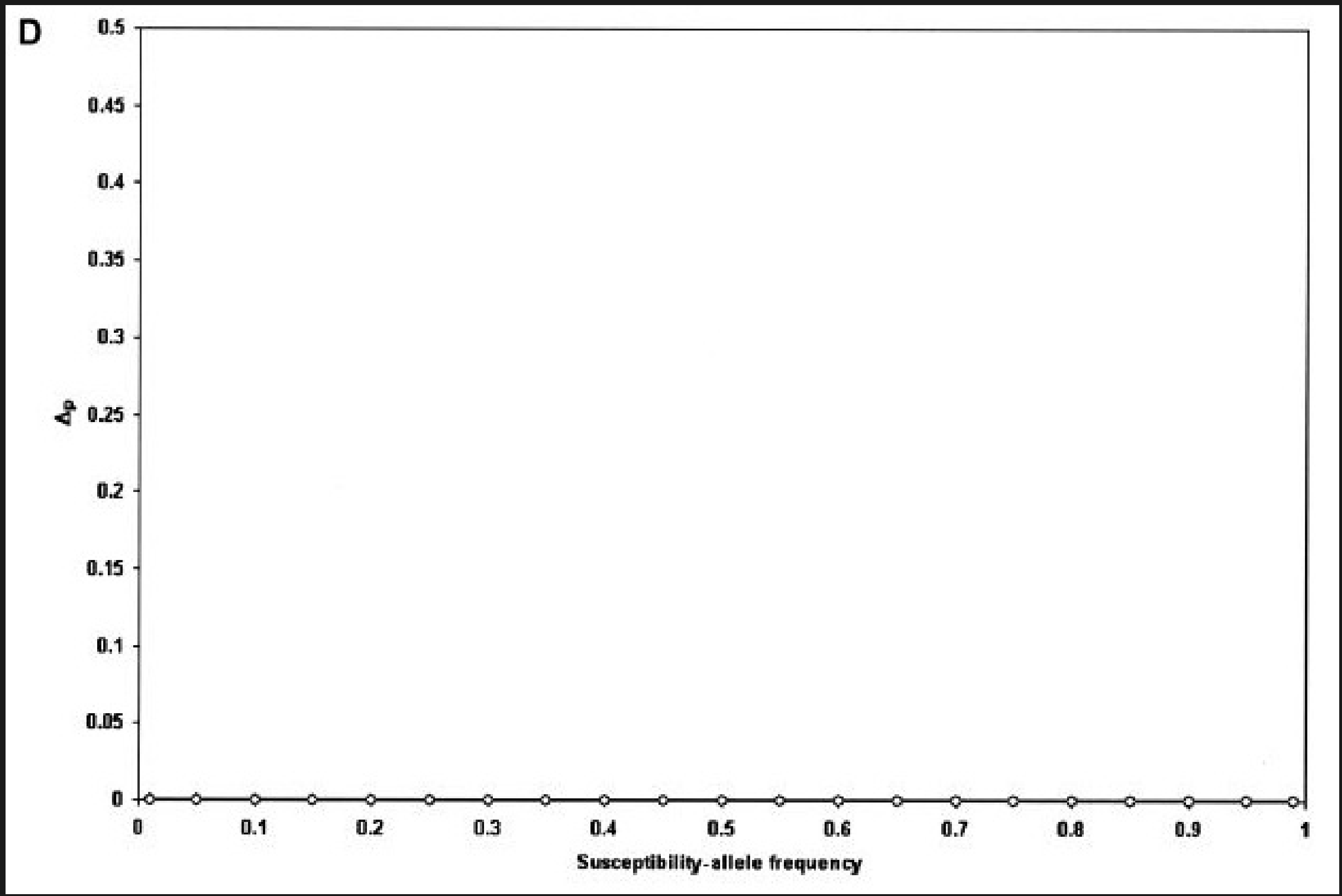


Epstein and Satten, Am J Hum Genet 73:1316-1329, 2003, Fig 1

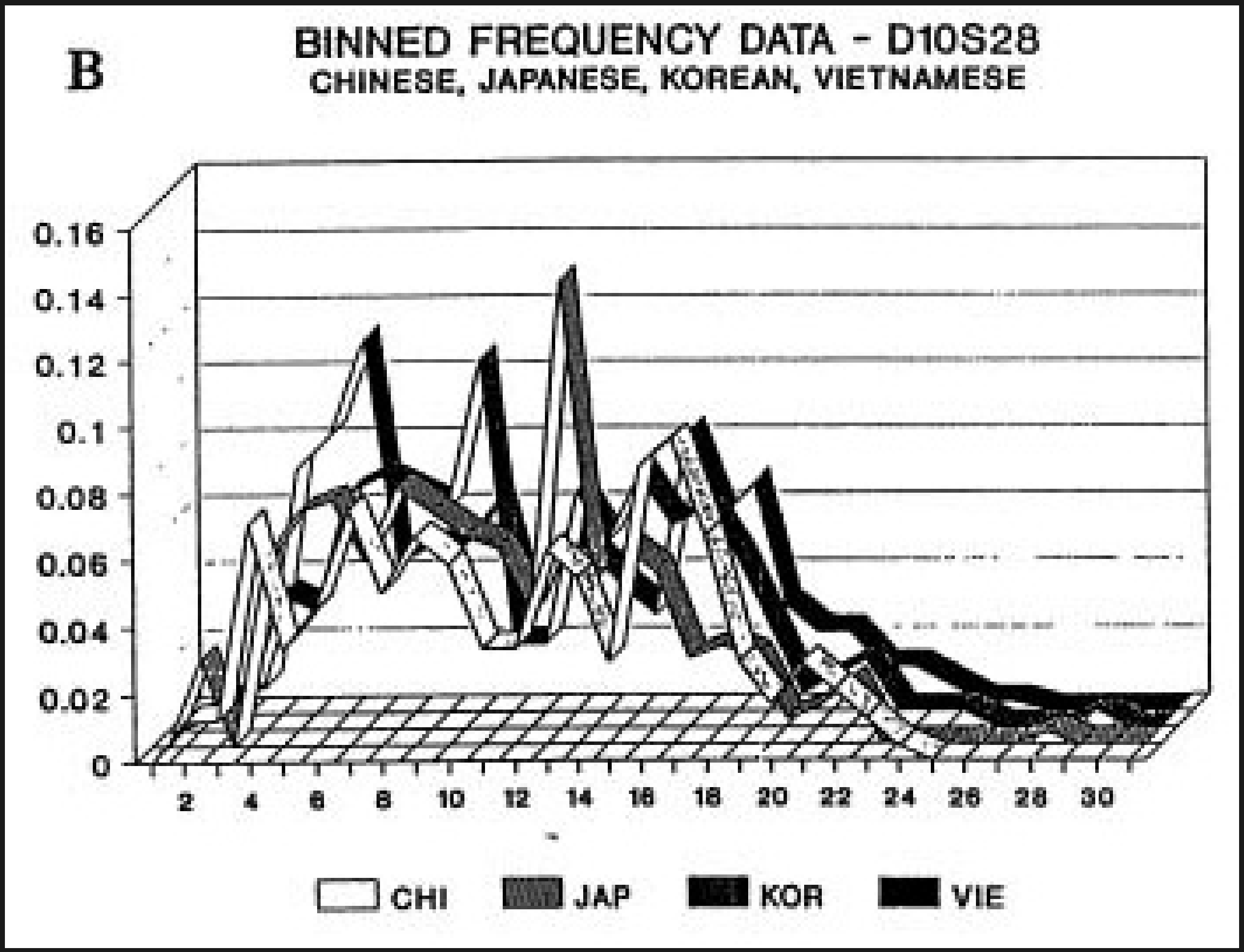


(a) (b)

Mykland et al., J Am Stat Asso 90:233-241, 1995, Fig 1



Wittke-Thompson et al., Am J Hum Genet 76:967-986, Fig 1



Roeder, Stat Sci 9:222-278, 1994, Fig 4