Principles of Data Visualization

Decoding

August 13, 2026

Ronald Geskus
Oxford University Clinical Research Unit
Centre for Tropical Medicine
Ho Chi Minh City, Viet Nam

Outline

Introduction

Principles of Data Visualisation
- Scale and range
- Theories of perception

Decoding: William Cleveland
- A hierarchy of visual decoding
  - Size (numeric)
  - Group (categorical)

Colour

Aesthetics: Tufte

Final words

Contents of this course

How to make an informative visual representation of data and/or results

- Part I: process of graph construction (encoding)
  - How to represent ("map") data/results in graph
  - The structure of a graph; the grammar of graphics
- Part II: principles of graph perception and interpretation (decoding)
  - What happens when you see and read a graph
  - Principles for creating truthful and informative graphs

Points to consider

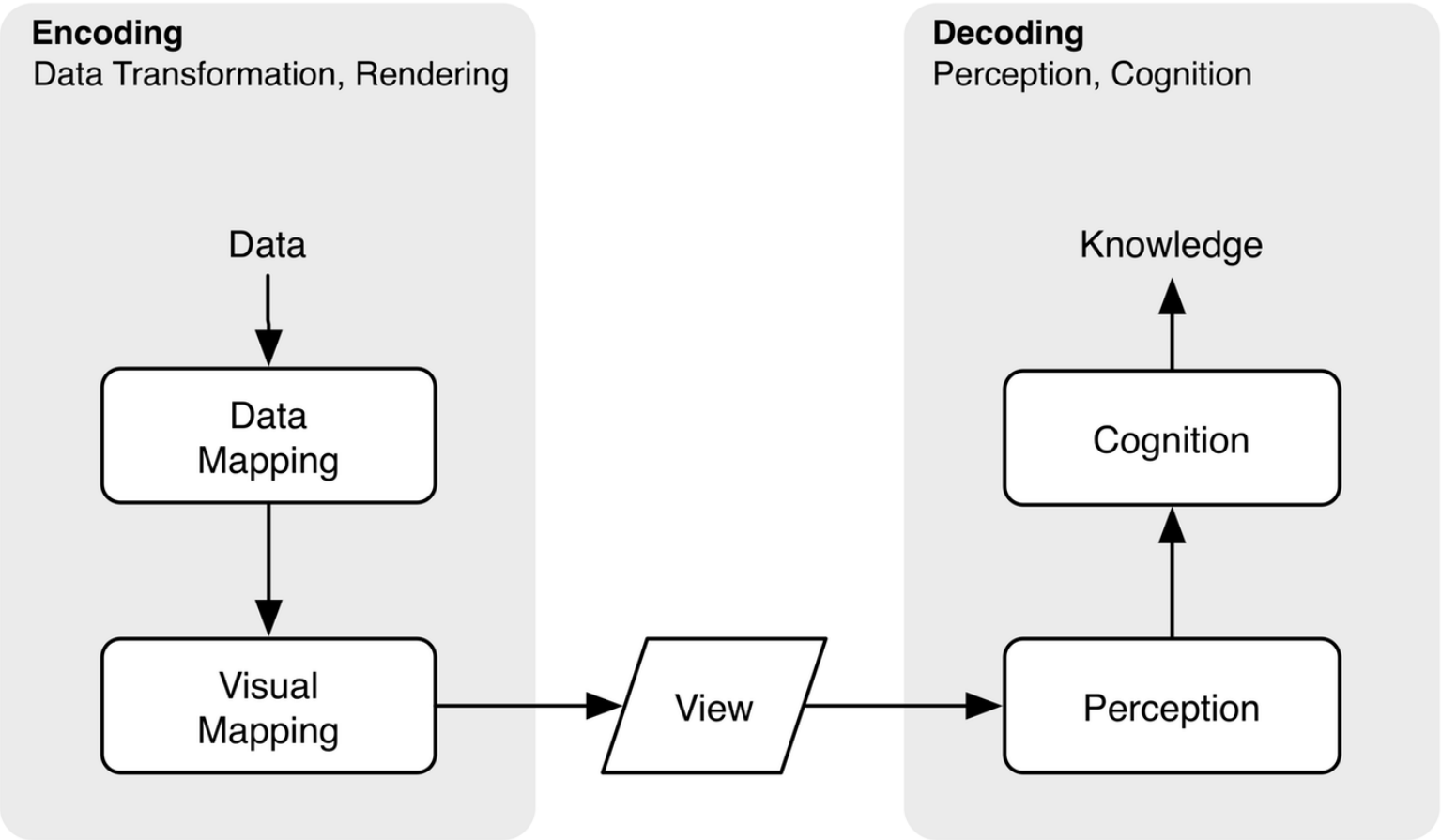
- Why do we make a graph? (purpose)
  1. We want to explore and inspect our data or results
  2. Present data or results, often with a message, to others
- Where do we show the graph: research paper, conference, website
- Who is the audience. Level of background knowledge
- What do we want to show (data)
  - Graph: visual representation of information ("encoding") mapping from data/results to figure
  - Show values in isolation and/or focus on patterns

Contents of this course

How to make an informative visual representation of data and/or results

- Part I: process of graph construction (encoding)
  - How to represent ("map") data/results in graph
  - The structure of a graph; the grammar of graphics
- Part II: principles of graph perception and interpretation (decoding)
  - What happens when you see and read a graph
  - Principles for creating truthful and informative graphs

Creating (mapping) and perceiving (seeing and reading)



Grammar of Graphics (R terminology)

Statistical Graph mapping from data to aesthetic attributes of geometric objects

Geometric Object what you see on the plot (points, lines, bars)

Aesthetic Attributes "role" of variables in plot: location (often along x-axis and y-axis), colour, size, shape
Scales map data values to aesthetic attributes.
Maybe with position adjustment along axis (stacked or dodged barplot, point jitter)

Statistical Transformation

Coordinate System Cartesian, polar, map projections. (Networks, trees have no coordinate system)

Faceting separate panels (subfigures)

Grammar of Graphics (R terminology)

Statistical Graph mapping from data to aesthetic attributes of geometric objects

Geometric Object what you see on the plot (points, lines, bars)

Aesthetic Attributes "role" of variables in plot: location (often along x-axis and y-axis), colour, size, shape
Scales map data values to aesthetic attributes.
Maybe with position adjustment along axis (stacked or dodged barplot, point jitter)

Statistical Transformation

Coordinate System Cartesian, polar, map projections.
(Networks, trees have no coordinate system)

Faceting separate panels (subfigures)

Not about graph type; ($\approx$ language, words).
Not about font size, background colour . . . Can be specified in a theme

Outline

Introduction

Principles of Data Visualisation
- Scale and range
- Theories of perception

Decoding: William Cleveland
- A hierarchy of visual decoding
  - Size (numeric)
  - Group (categorical)

Colour

Aesthetics: Tufte

Final words

Criteria for graph quality

- Truthful
  - based on data collected in honest research
  - objective display of information
  - use appropriate scale and range

Criteria for graph quality

- Truthful
  - based on data collected in honest research
  - objective display of information
  - use appropriate scale and range

Outline

Introduction

Principles of Data Visualisation
- Scale and range
- Theories of perception

Decoding: William Cleveland
- A hierarchy of visual decoding
  - Size (numeric)
  - Group (categorical)

Colour

Aesthetics: Tufte

Final words

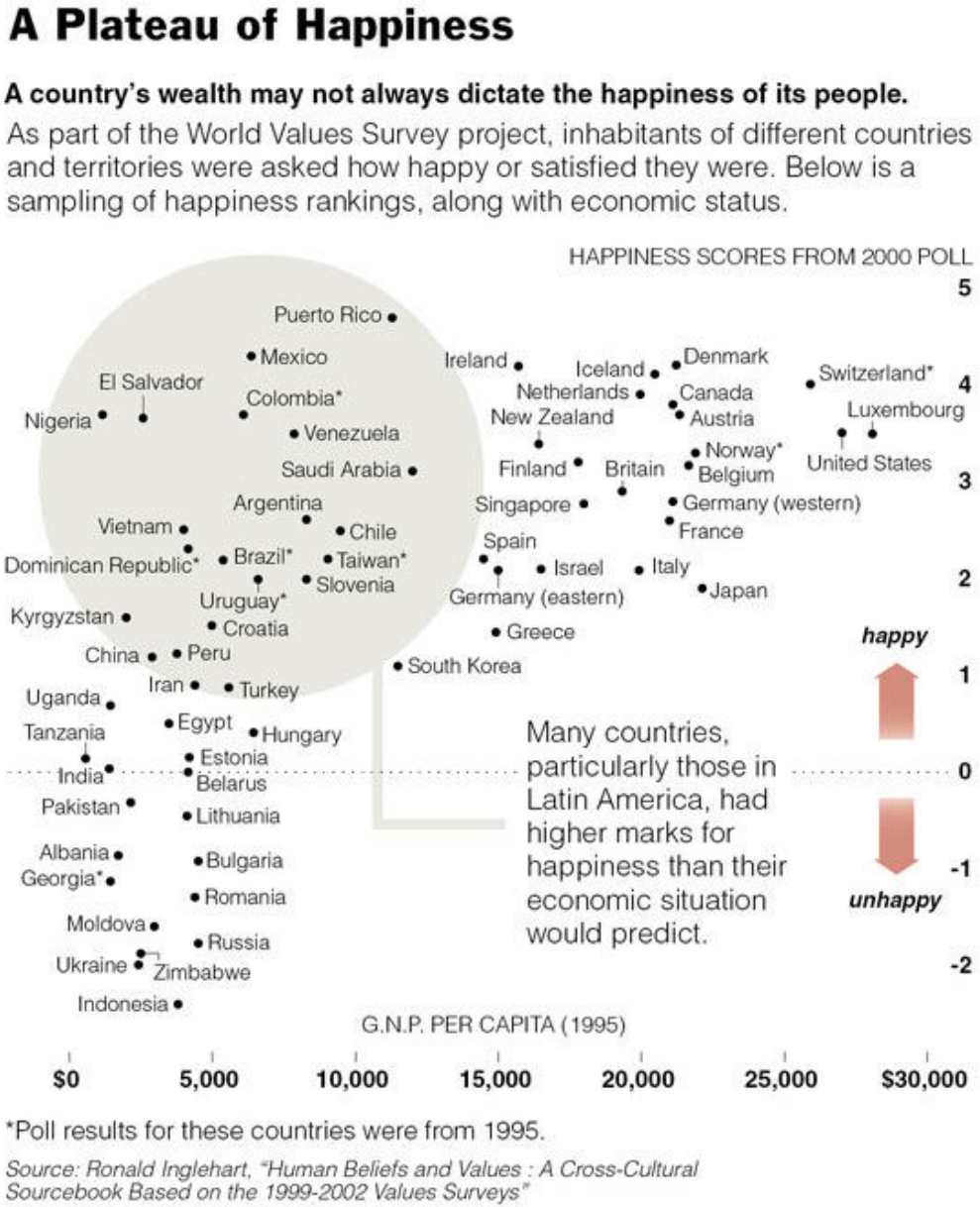
Scales on the axes, some principles

- Linear scale: values as measured
  - Most easy interpretation of values

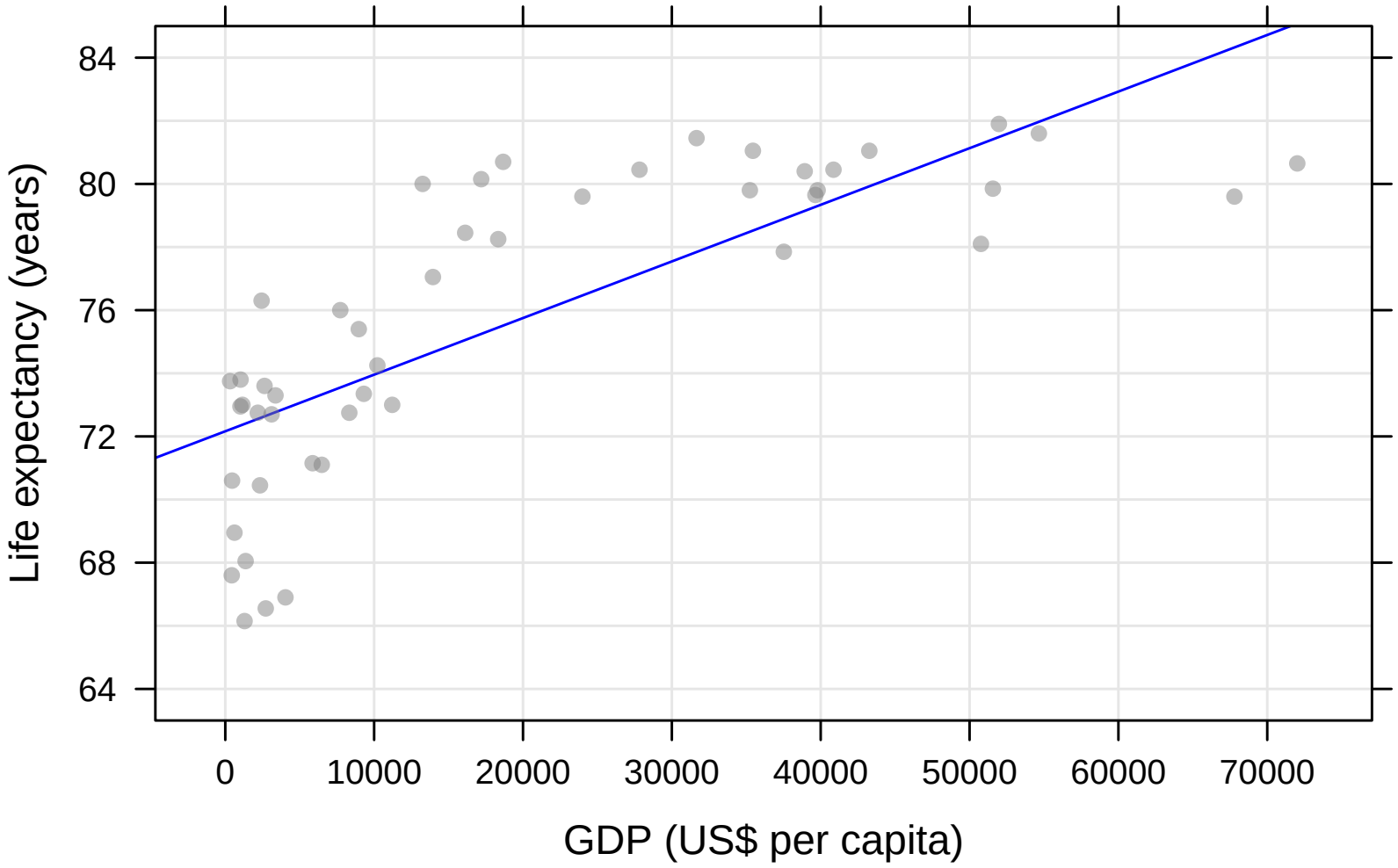
  But:
  - Distribution may be very skewed
  - May give wrong impression of relationship

OTN Mombasa 2025

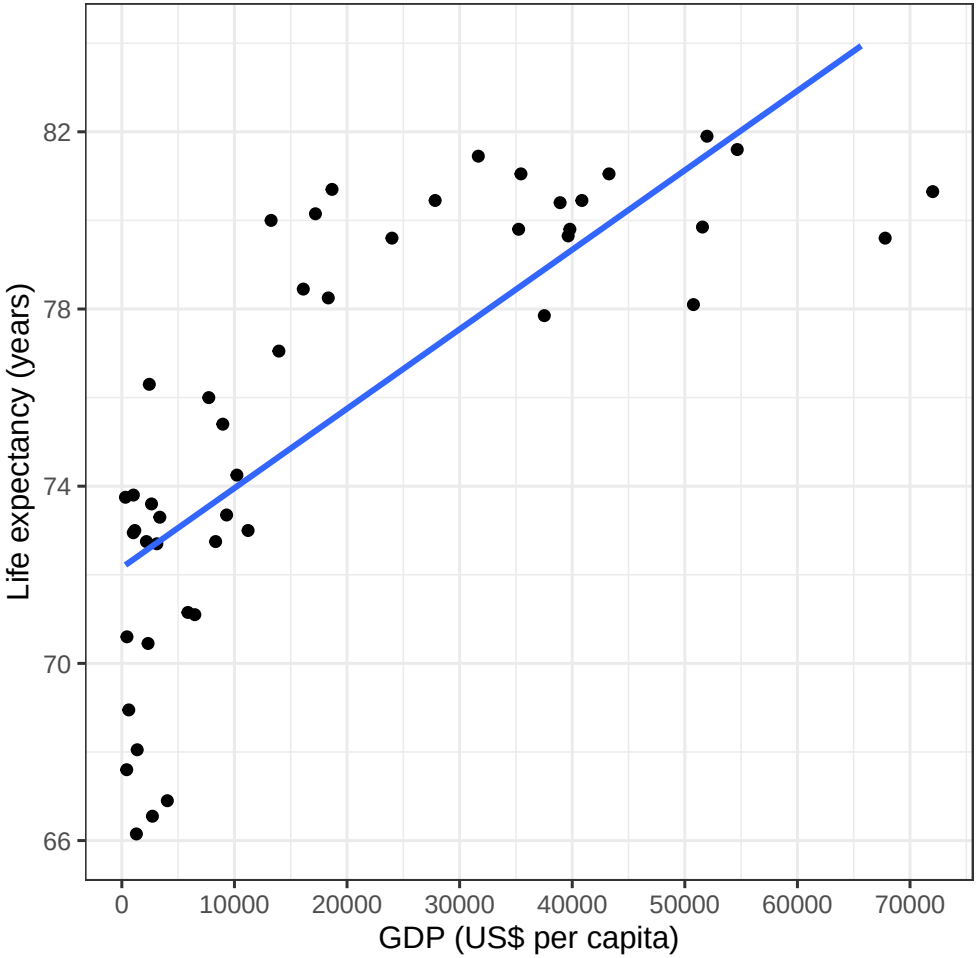
Choice of scale influences interpretation

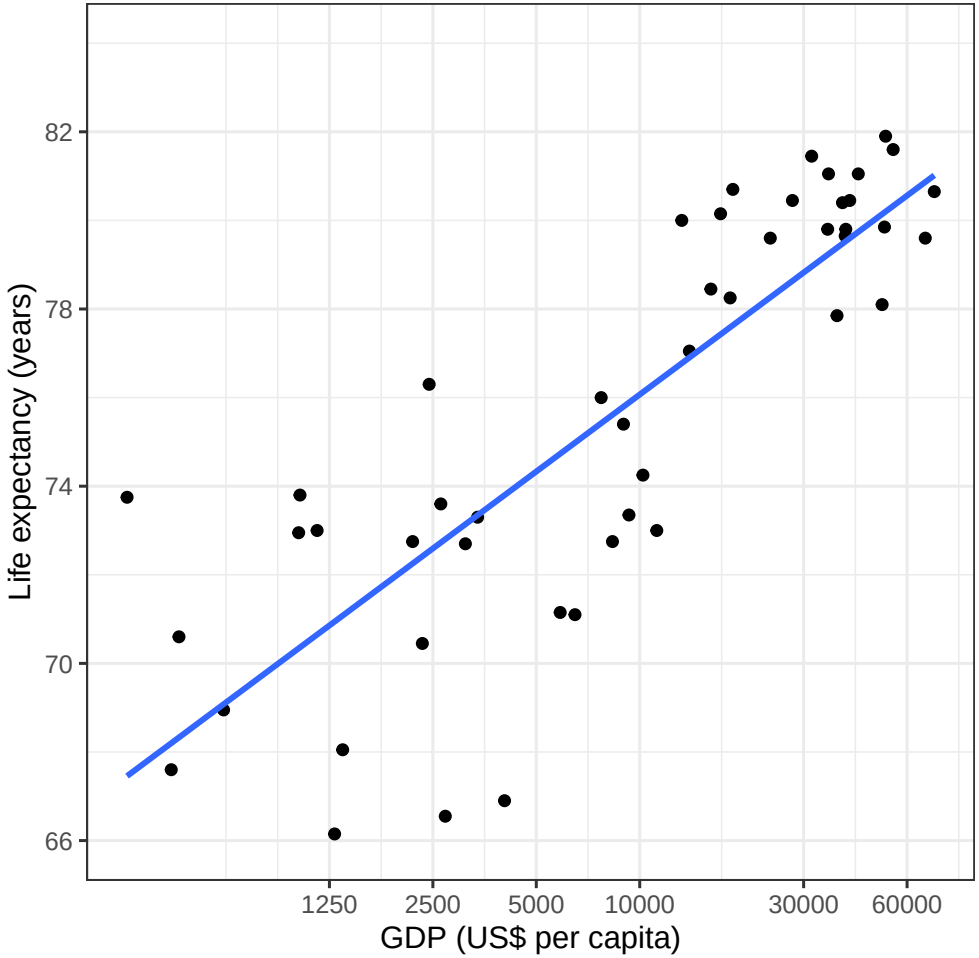


Example: WHO European Health Report, 2007



WHO data 2007, two scales



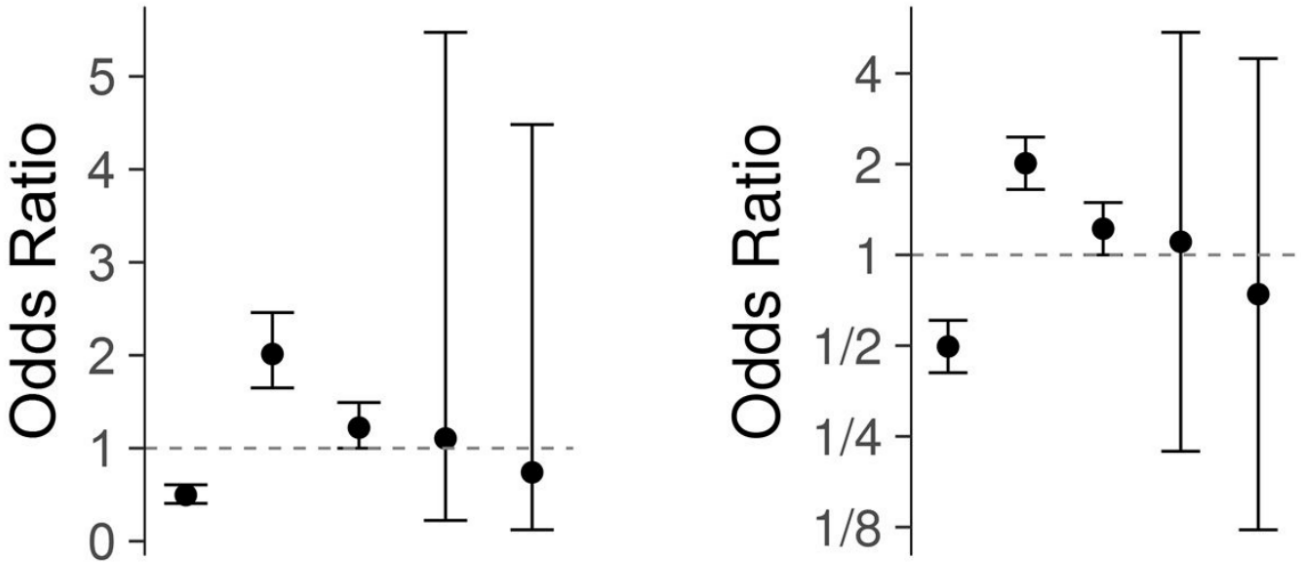


Scales on the axes, some principles

- Linear scale: values as measured
  - Most easy interpretation of values

  But:
  - Distribution may be very skewed
  - May give wrong impression of relationship
- Log scale
  - Distance reflects percentage change:
    $\log(x_2) - \log(x_1) = \log(x_2/x_1)$
    e.g. $\log_2(2500) - \log_2(1250) = \log_2(2) = 1$ unit on scale
  - Can still use original values as labels
  - Always use log scale for ratios

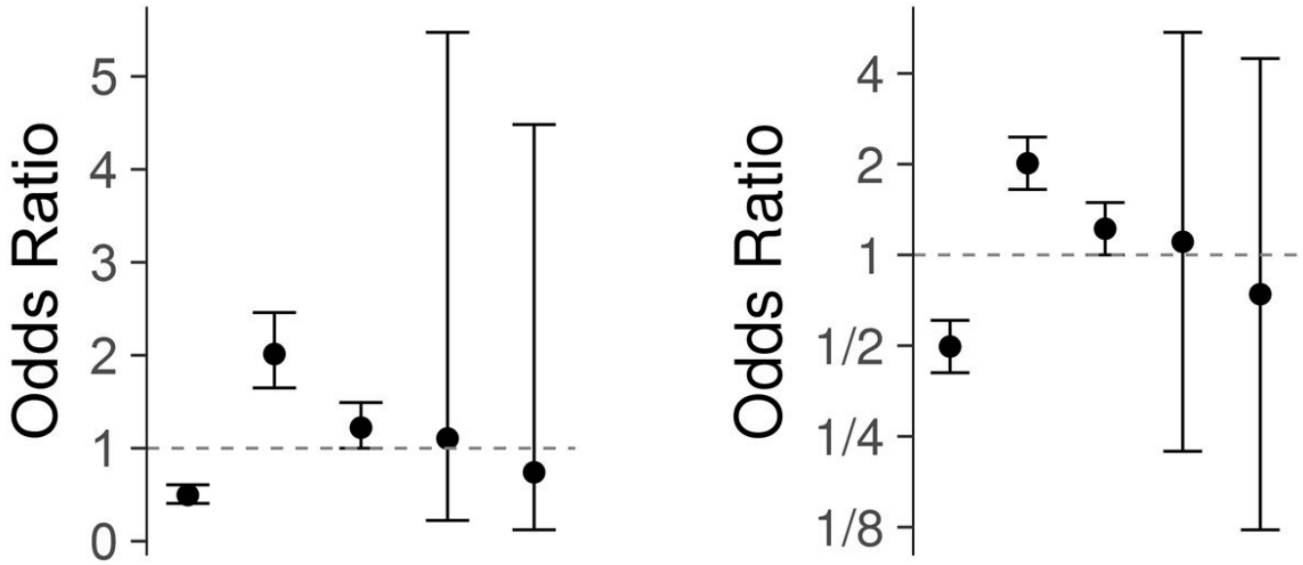
Always use a log scale for ratios



$\left(P\text{: probability outcome; odds} = \frac{P}{1-P}\text{;}\right.$

$\left.\text{odds ratio odds}(A)/\text{odds}(B) = \frac{P_A}{1-P_A} / \frac{P_B}{1-P_B}\right)$

Always use a log scale for ratios

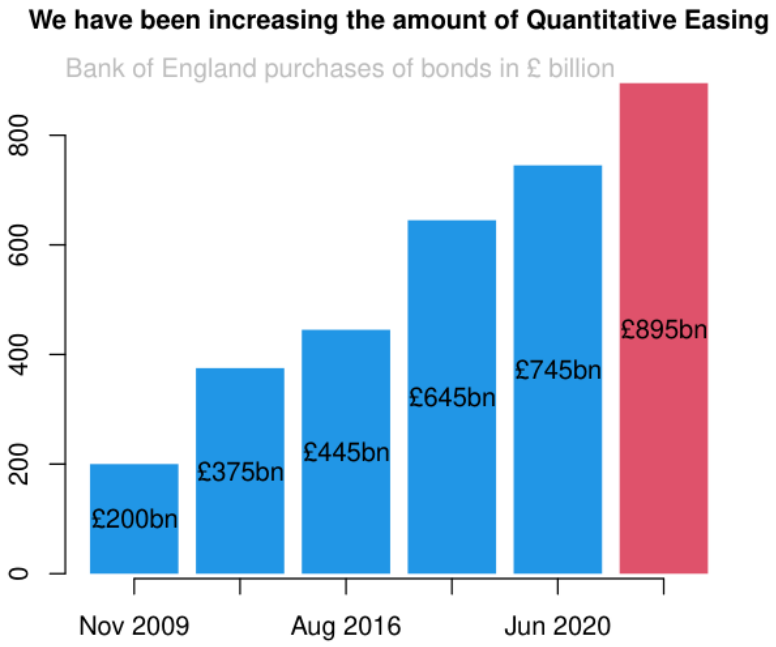


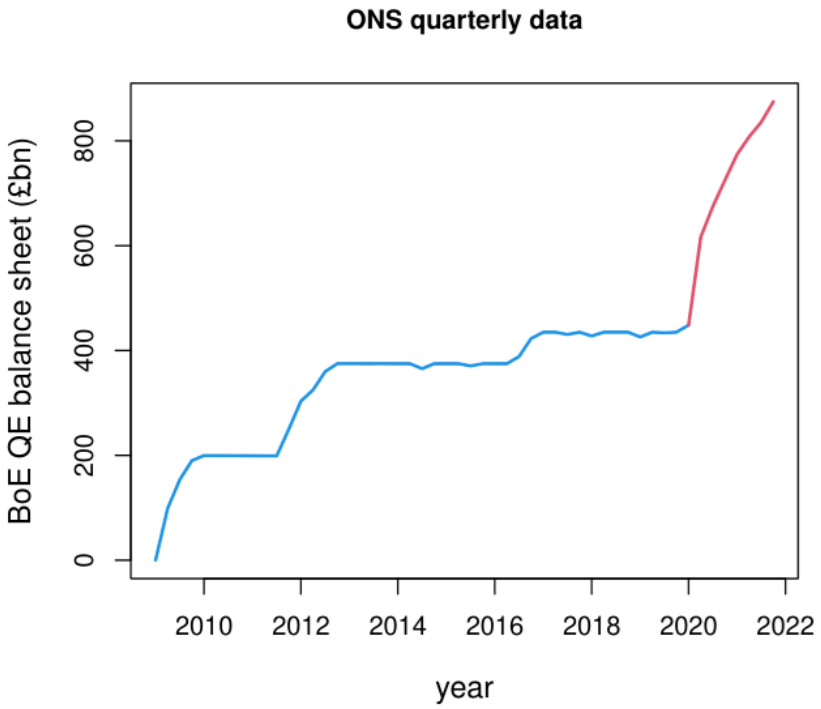
$\left(P\text{: probability outcome; odds} = \frac{P}{1-P}\text{;}\right.$

$\left.\text{odds ratio odds}(A)/\text{odds}(B) = \frac{P_A}{1-P_A} / \frac{P_B}{1-P_B}\right)$

$$\text{If OR(F/M)} = \frac{\text{odds(F)}}{\text{odds(M)}} = 2\text{, then OR(M/F)} = \frac{\text{odds(M)}}{\text{odds(F)}} = 0.5$$

Change in Quantitative Easing



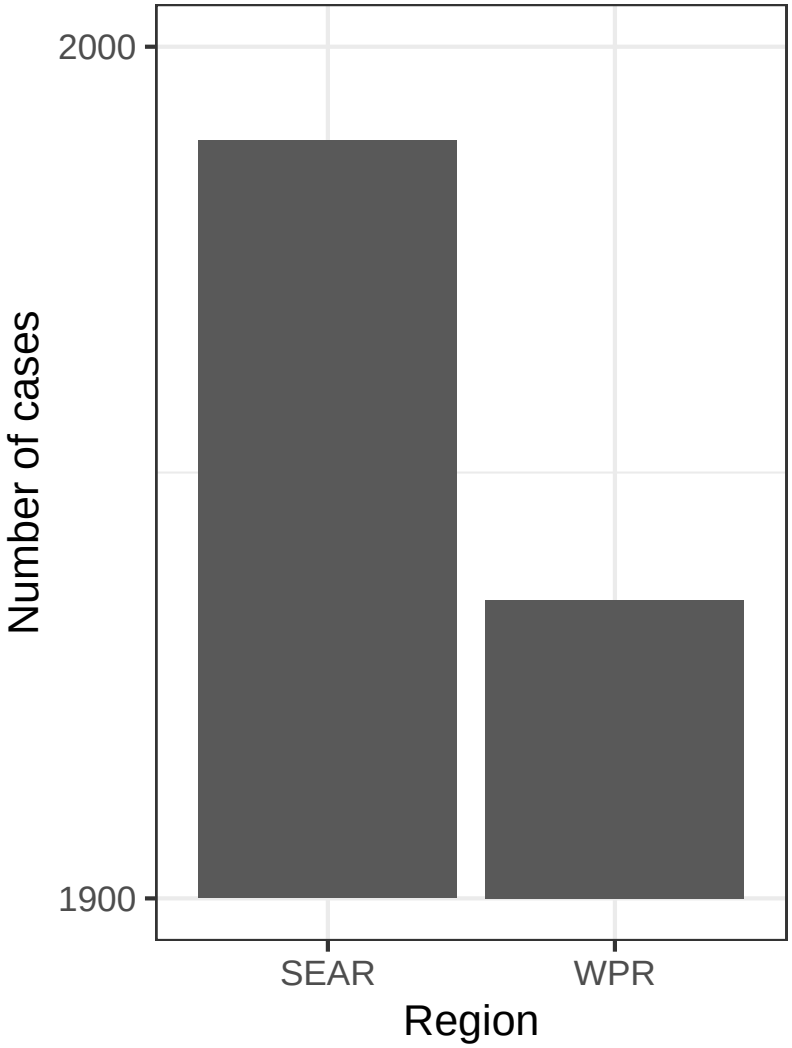


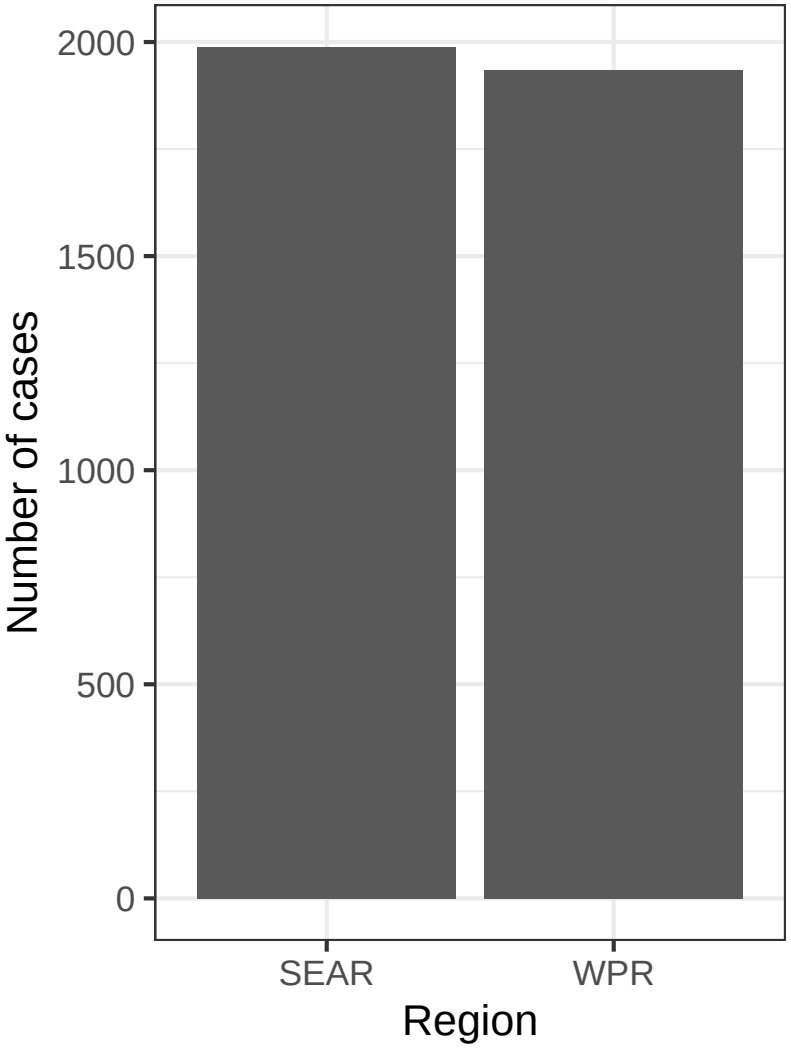
Scales on the axes, some principles

- Linear scale: values as measured
  - Most easy interpretation of values

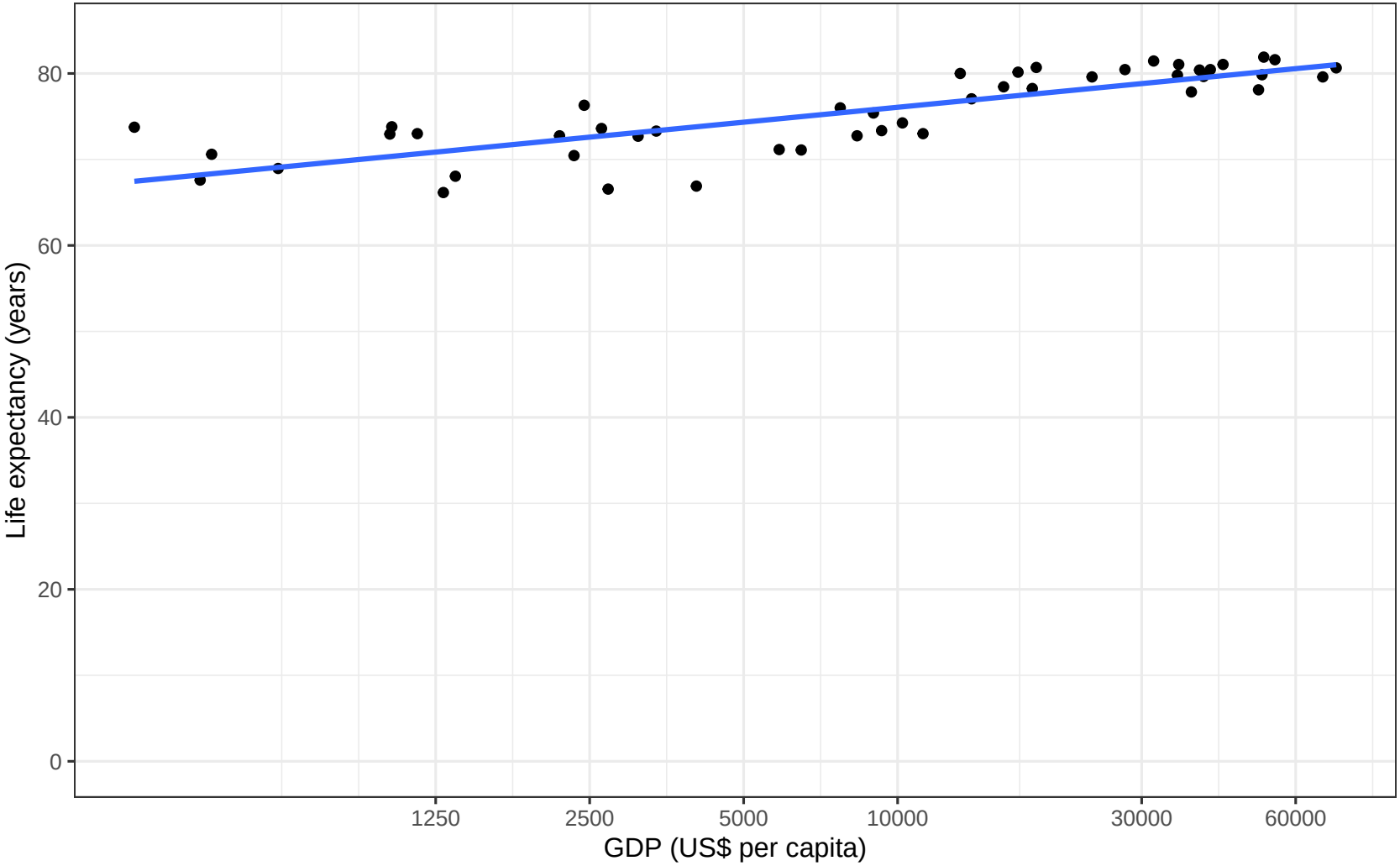
  But:
  - Distribution may be very skewed
  - May give wrong impression of relationship
- Log scale
  - Distance reflects percentage change:
    $\log(x_2) - \log(x_1) = \log(x_2/x_1)$
    e.g. $\log_2(2500) - \log_2(1250) = \log_2(2) = 1$ unit on scale
  - Can still use original values as labels
  - Always use log scale for ratios
- Range: is zero included?
  Does zero have a meaning?

Include zero: measles cases in October 2019, two WHO regions





Don't include zero: no country has life expectancy $<$ 60

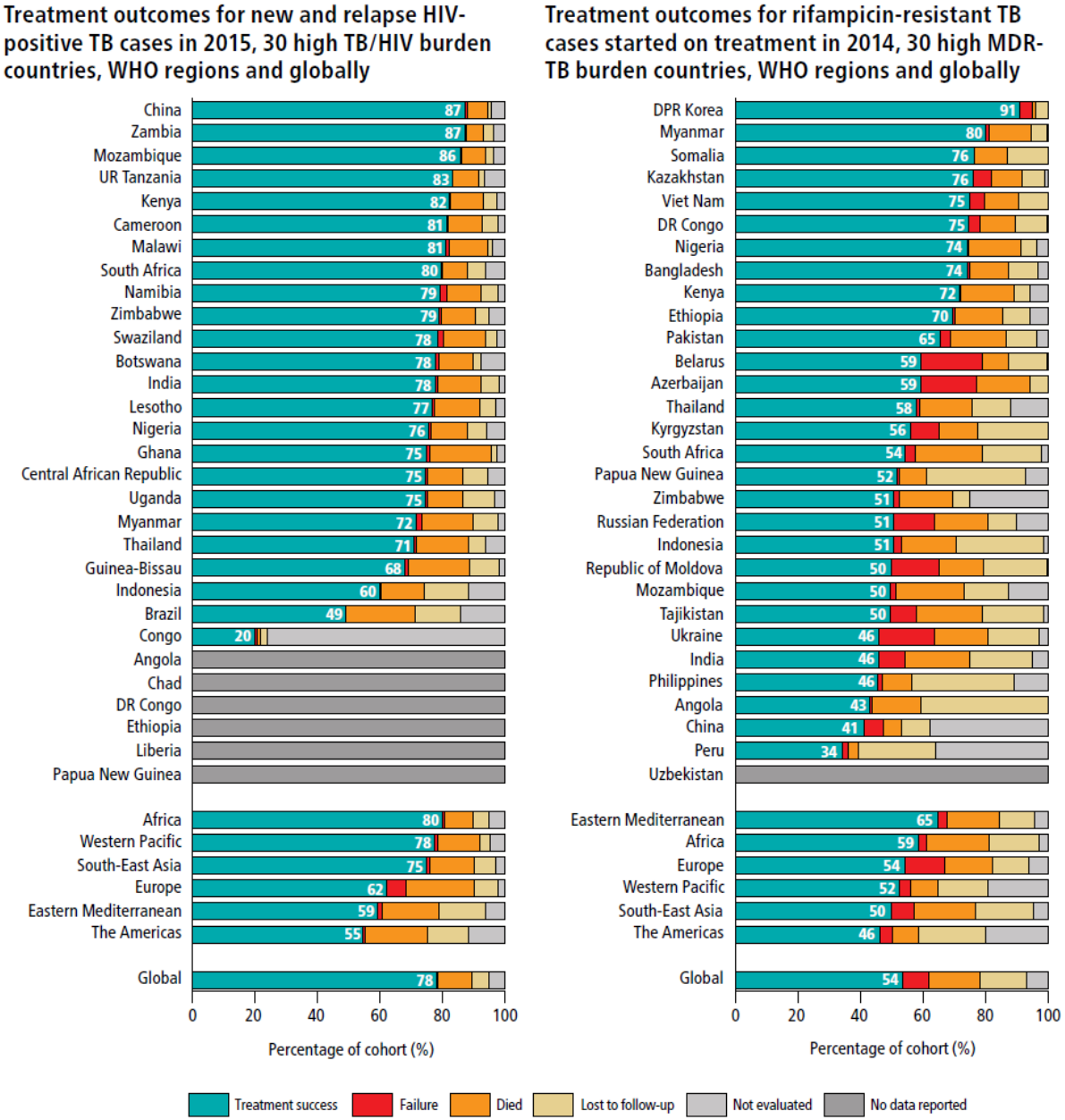


Scales on the axes, some principles

- Linear scale: values as measured
  - Most easy interpretation of values

  But:
  - Distribution may be very skewed
  - May give wrong impression of relationship
- Log scale
  - Distance reflects percentage change:
    $\log(x_2) - \log(x_1) = \log(x_2/x_1)$
    e.g. $\log_2(2500) - \log_2(1250) = \log_2(2) = 1$ unit on scale
  - Can still use original values as labels
  - Always use log scale for ratios
- Range: is zero included?
  Does zero have a meaning?
- Categorical data: ordering by i) alphabet? ii) value? iii) subgroup? (e.g. disease type, region)

WHO Global TB Report 2017



Criteria for graph quality

- Truthful
  - based on data collected in honest research
  - objective display of information
  - use appropriate scale and range
- Intuitive: easy to understand, concise without oversimplifying (see Section “Aesthetics: Tufte”)

Criteria for graph quality

- Truthful
  - based on data collected in honest research
  - objective display of information
  - use appropriate scale and range
- Intuitive: easy to understand, concise without oversimplifying (see Section “Aesthetics: Tufte”)
- Informative (see Section “Decoding: William Cleveland”)
  - use appropriate graph type
  - use proximity and alignment to facilitate comparison
  - use labels and annotations to add clarity to the message

Criteria for graph quality

- Truthful
  - based on data collected in honest research
  - objective display of information
  - use appropriate scale and range
- Intuitive: easy to understand, concise without oversimplifying (see Section “Aesthetics: Tufte”)
- Informative (see Section “Decoding: William Cleveland”)
  - use appropriate graph type
  - use proximity and alignment to facilitate comparison
  - use labels and annotations to add clarity to the message
- Insightful: reveals evidence hard to observe otherwise (see Section “Aesthetics: Tufte”)

Criteria for graph quality

- Truthful
  - based on data collected in honest research
  - objective display of information
  - use appropriate scale and range
- Intuitive: easy to understand, concise without oversimplifying (see Section “Aesthetics: Tufte”)
- Informative (see Section “Decoding: William Cleveland”)
  - use appropriate graph type
  - use proximity and alignment to facilitate comparison
  - use labels and annotations to add clarity to the message
- Insightful: reveals evidence hard to observe otherwise (see Section “Aesthetics: Tufte”)
- Visually appealing; but honesty and clarity come first

Outline

Introduction

Principles of Data Visualisation
- Scale and range
- Theories of perception

Decoding: William Cleveland
- A hierarchy of visual decoding
  - Size (numeric)
  - Group (categorical)

Colour

Aesthetics: Tufte

Final words

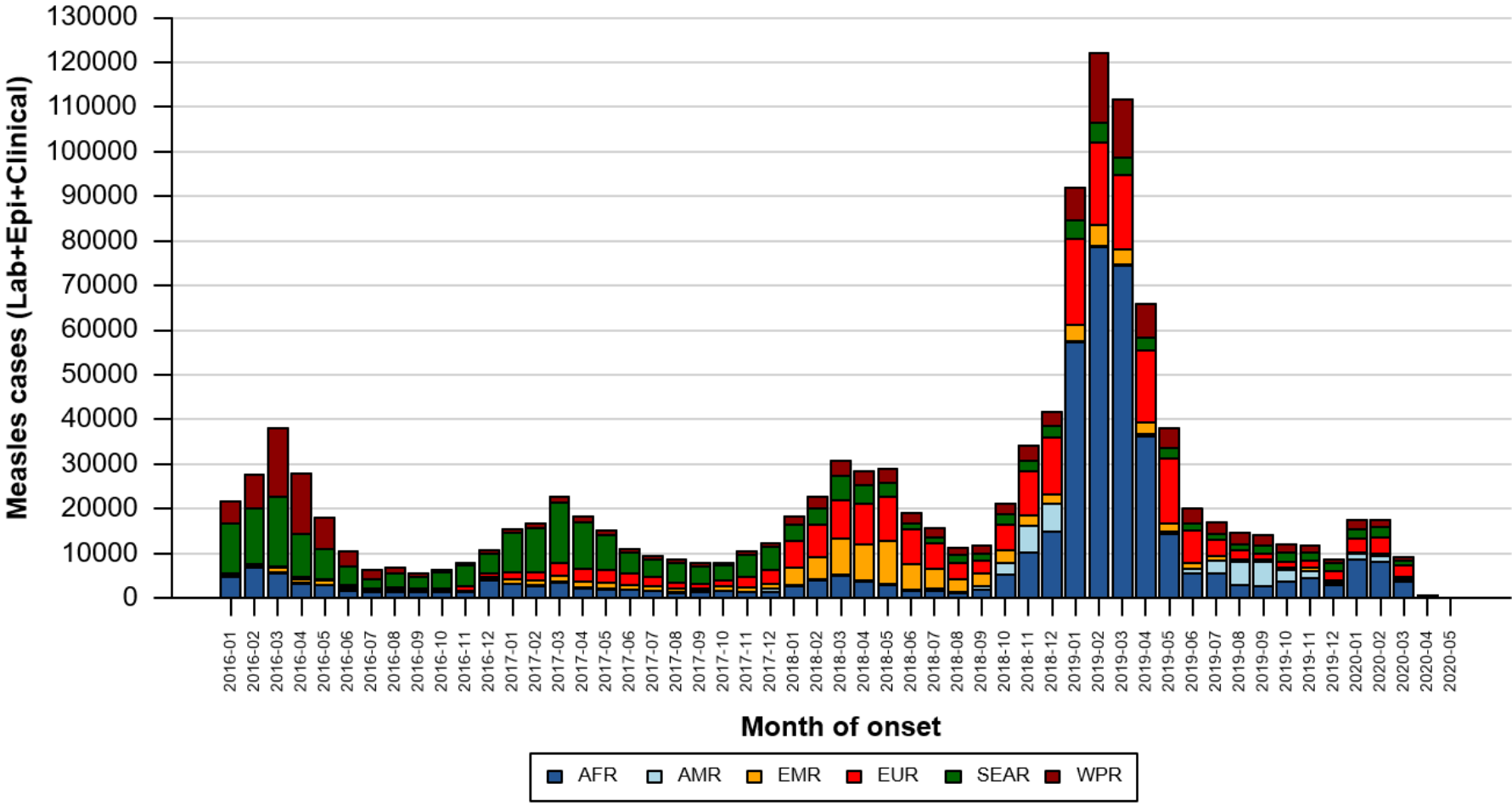
Two modes of perception: Seeing and Reading

- Daniel Kahneman: Thinking, Fast and Slow
  Fast: intuitive, “seeing”
  Slow: logical, “reading”

Two modes of perception: Seeing and Reading

- Daniel Kahneman: Thinking, Fast and Slow
  Fast: intuitive, “seeing”
  Slow: logical, “reading”
- What happens when you interpret a graph

Measles case distribution by month and WHO Region (2016-2020)



Two modes of perception: Seeing and Reading

- Daniel Kahneman: Thinking, Fast and Slow
  Fast: intuitive, “seeing”
  Slow: logical, “reading”
- What happens when you interpret a graph
  1. Graphical perception: first impression, see.
     “There was a peak in measles cases in the first half of 2019”
  2. Graphical cognition: draw conclusions, read.
     “The outbreak was mainly in the African region”
     “The number of cases in the South-East Asian region has declined over time”

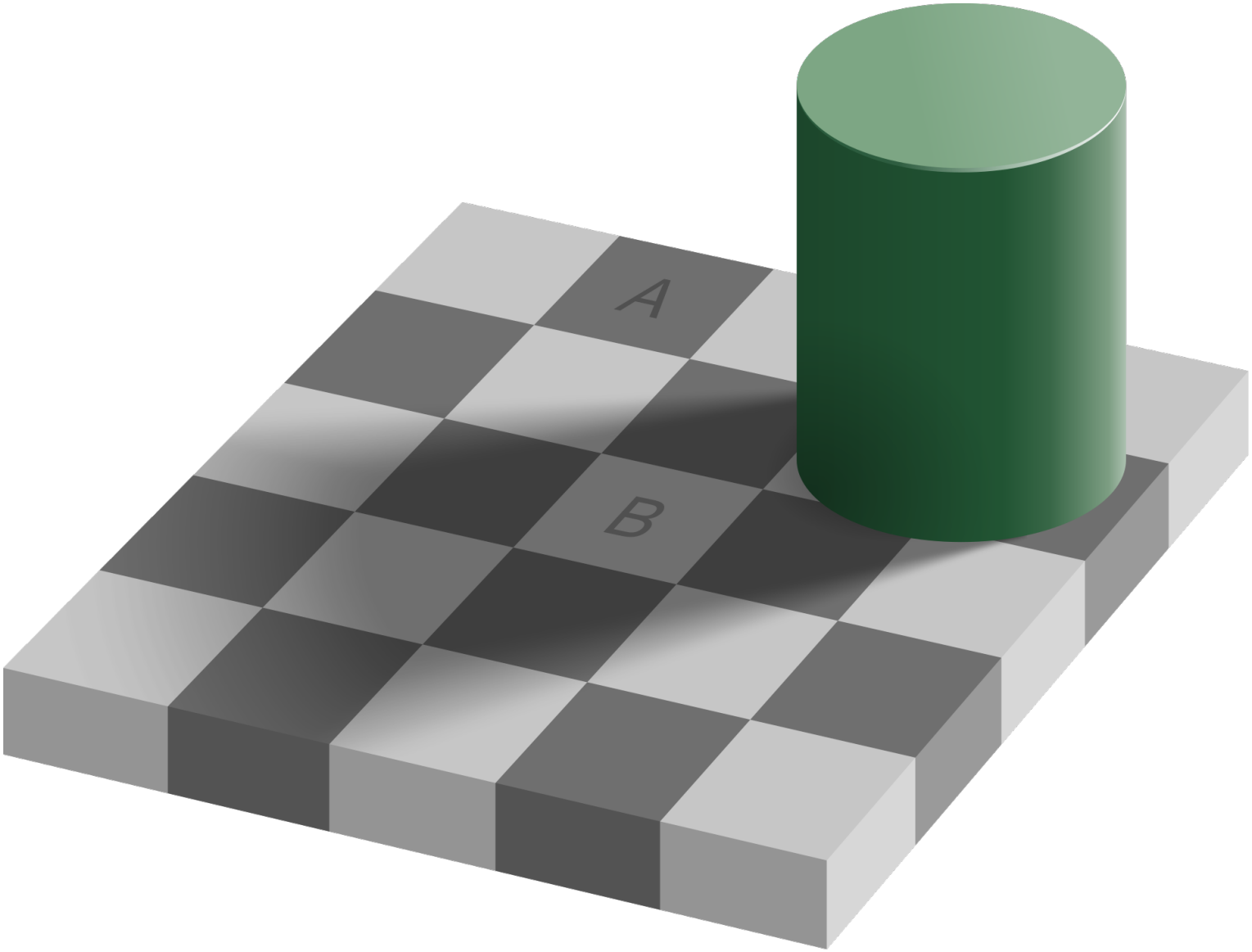
Human perception

- Gestalt principles. See e.g. https://www.toptal.com/designers/ui/gestalt-principles-of-design

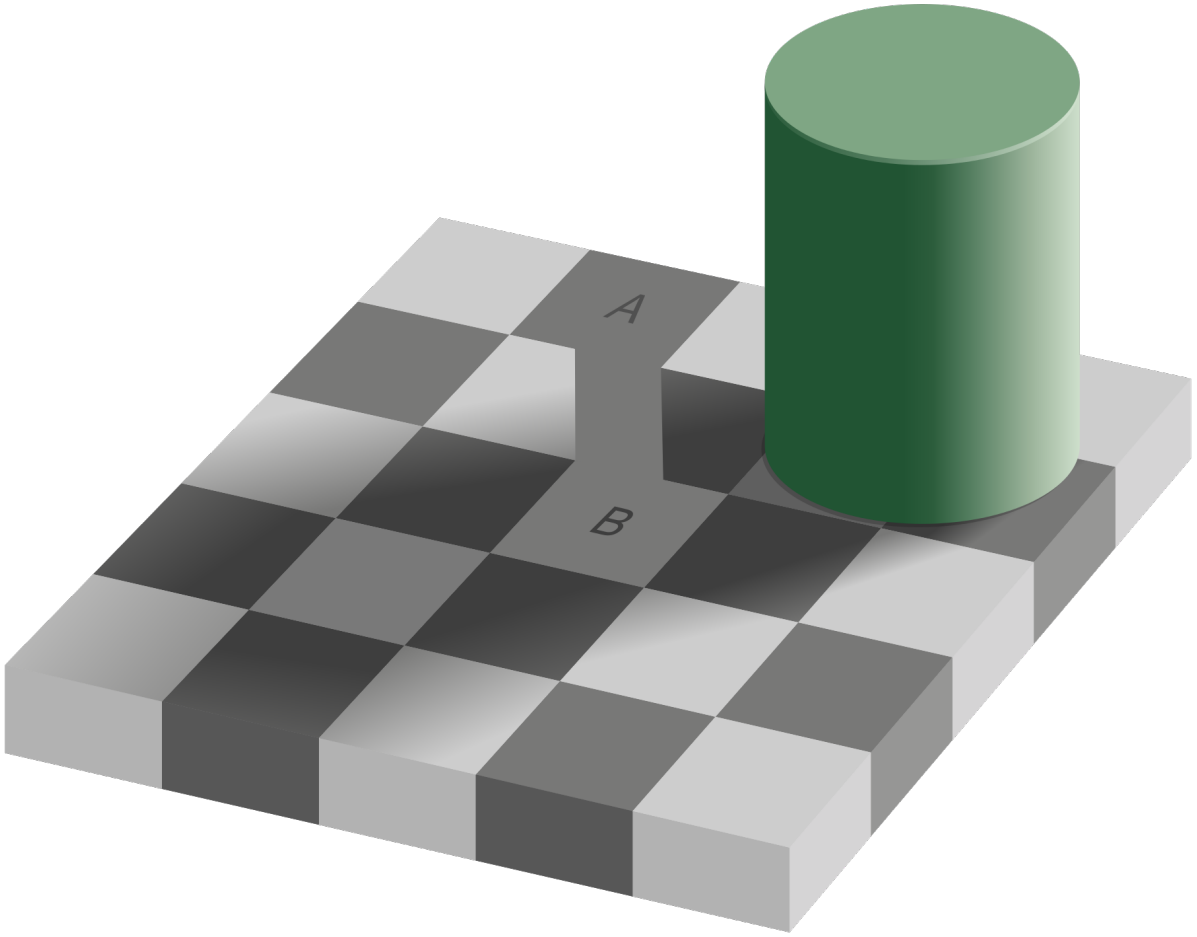
Human perception

- Gestalt principles. See e.g. https://www.toptal.com/designers/ui/gestalt-principles-of-design
- Anil Seth: Being You, A New Science of Consciousness (2021)
  - Perception: sensory signals that are interpreted based on the brain's expectations or beliefs about their causes.
  - We never experience sensory signals themselves, we only experience interpretations of them, which are constructions of the brain
  - Example: white paper remains white when you move from outdoors to indoors, because your brain knows its whiteness cannot change

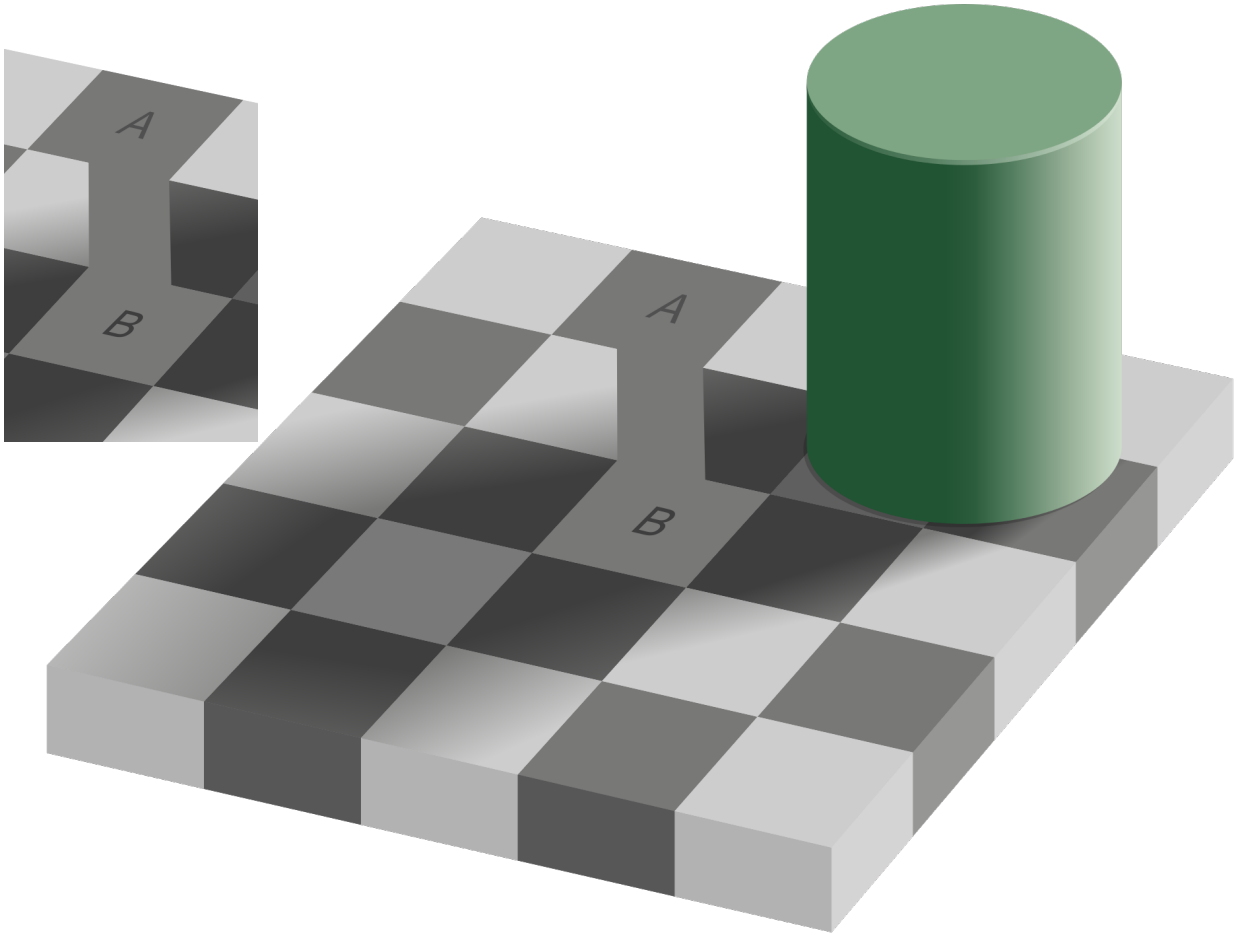
Adelson's Checkerboard



Adelson's Checkerboard (II)



Adelson's Checkerboard (II)



Criteria for graph quality

- Truthful
  - based on data collected in honest research
  - objective display of information
  - use appropriate scale and range
- Intuitive: easy to understand, concise without oversimplifying (see Section “Aesthetics: Tufte”)
- Informative (see Section “Decoding: William Cleveland”)
  - use appropriate graph type
  - use proximity and alignment to facilitate comparison
  - use labels and annotations to add clarity to the message
- Insightful: reveals evidence hard to observe otherwise (see Section “Aesthetics: Tufte”)
- Visually appealing; but honesty and clarity come first

Outline

Introduction

Principles of Data Visualisation
- Scale and range
- Theories of perception

Decoding: William Cleveland
- A hierarchy of visual decoding
  - Size (numeric)
  - Group (categorical)

Colour

Aesthetics: Tufte

Final words

William Cleveland: The Elements of Graphing Data

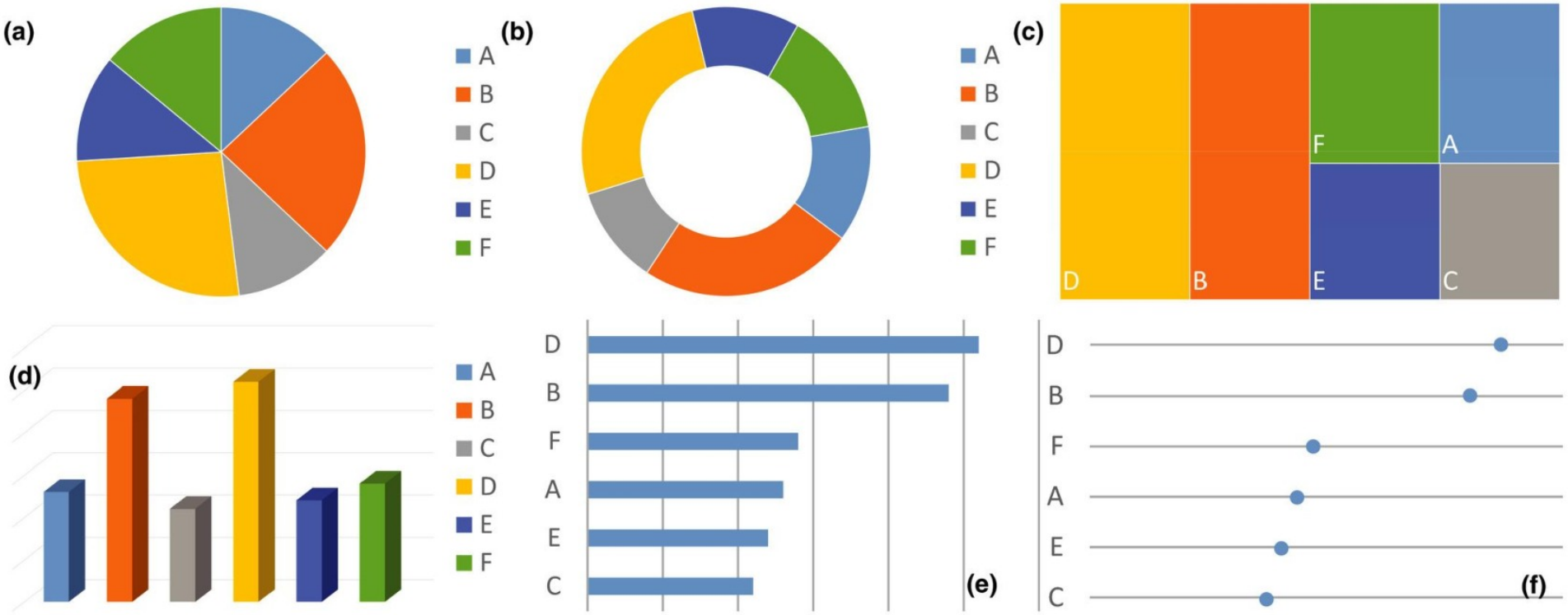
- Which graph is better at accurately revealing relevant information
- Study on most efficient ways to transfer information from data to a graph.
  Guidelines on creating informative graphs

William Cleveland: The Elements of Graphing Data

- Which graph is better at accurately revealing relevant information
- Study on most efficient ways to transfer information from data to a graph.
  Guidelines on creating informative graphs
- Pattern recognition. Assessment of relative magnitude of values: detection, assembly and estimation

detection	recognition of aesthetic attributes (encoding of values via scales)
assembly	discerning patterns; grouping
estimation	assessment of relative magnitude of values ("how much more?")

How to visualize six numbers



Outline

Introduction

Principles of Data Visualisation
- Scale and range
- Theories of perception

Decoding: William Cleveland
- A hierarchy of visual decoding
  - Size (numeric)
  - Group (categorical)

Colour

Aesthetics: Tufte

Final words

Decoding

- Table look-up: attach value to what is shown; pattern recognition

Decoding

- Table look-up: attach value to what is shown; pattern recognition
- Size (numeric)

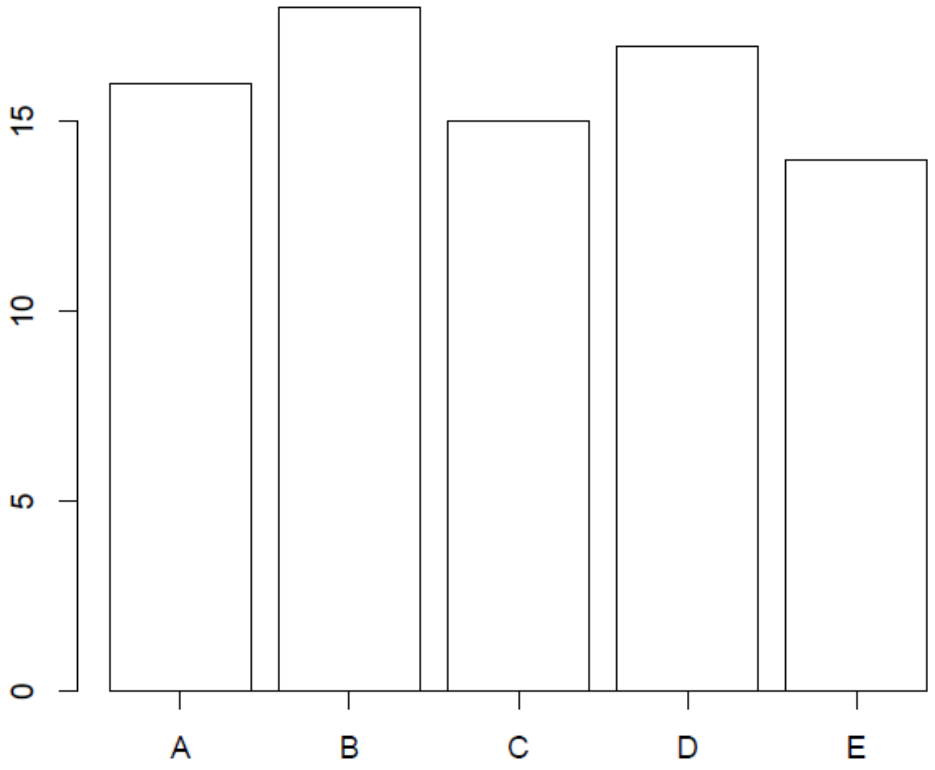
- Group (categorical)

Decoding

- Table look-up: attach value to what is shown; pattern recognition
- Size (numeric)
  In decreasing level of efficiency:
  1. Position along common scale

- Group (categorical)

1. Position along common scale

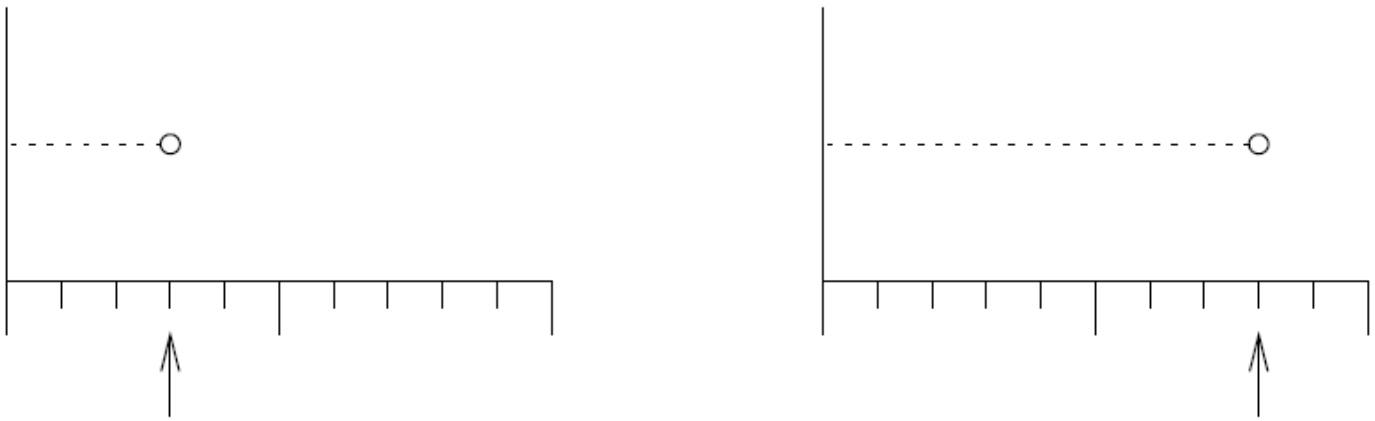


Decoding

- Table look-up: attach value to what is shown; pattern recognition
- Size (numeric)
  In decreasing level of efficiency:
  1. Position along common scale
  2. Position along identical, nonaligned scales

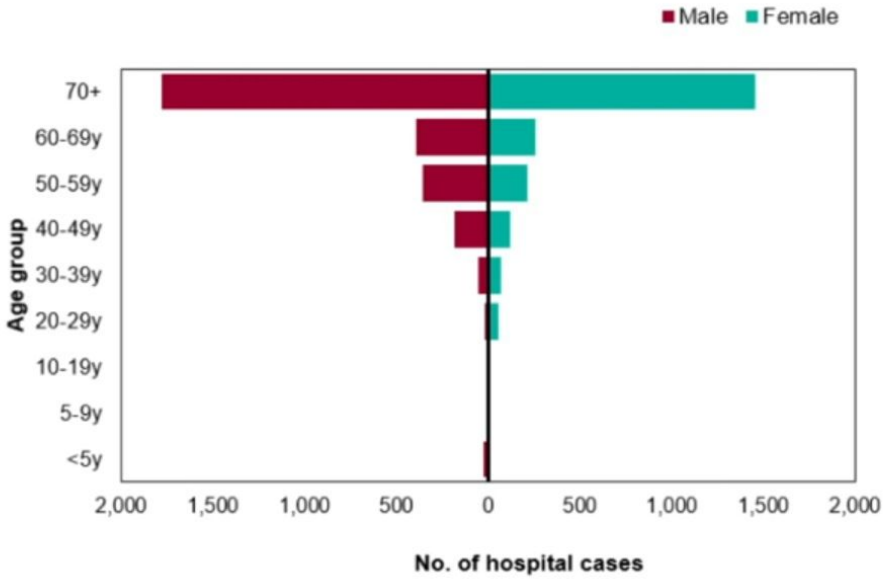
- Group (categorical)

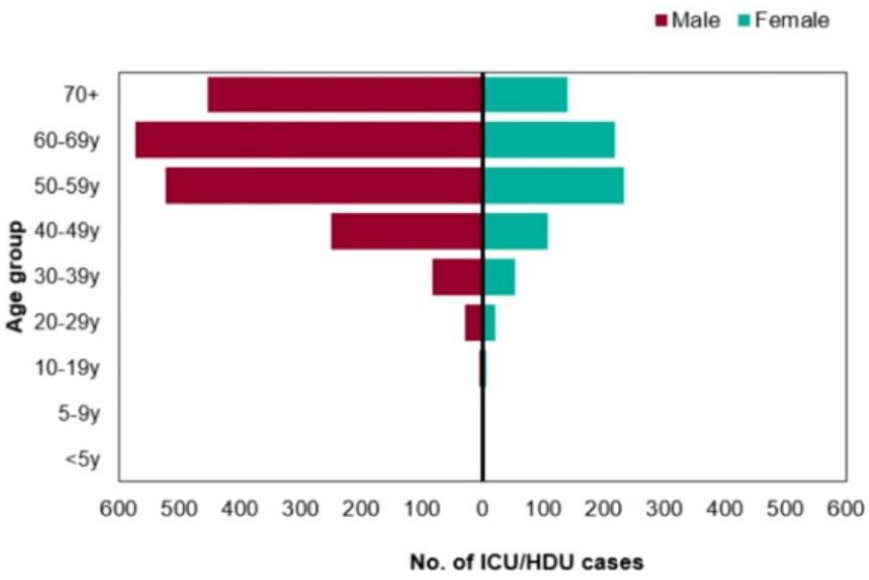
2. Position along identical, nonaligned scales



ICU mortality in covid-19

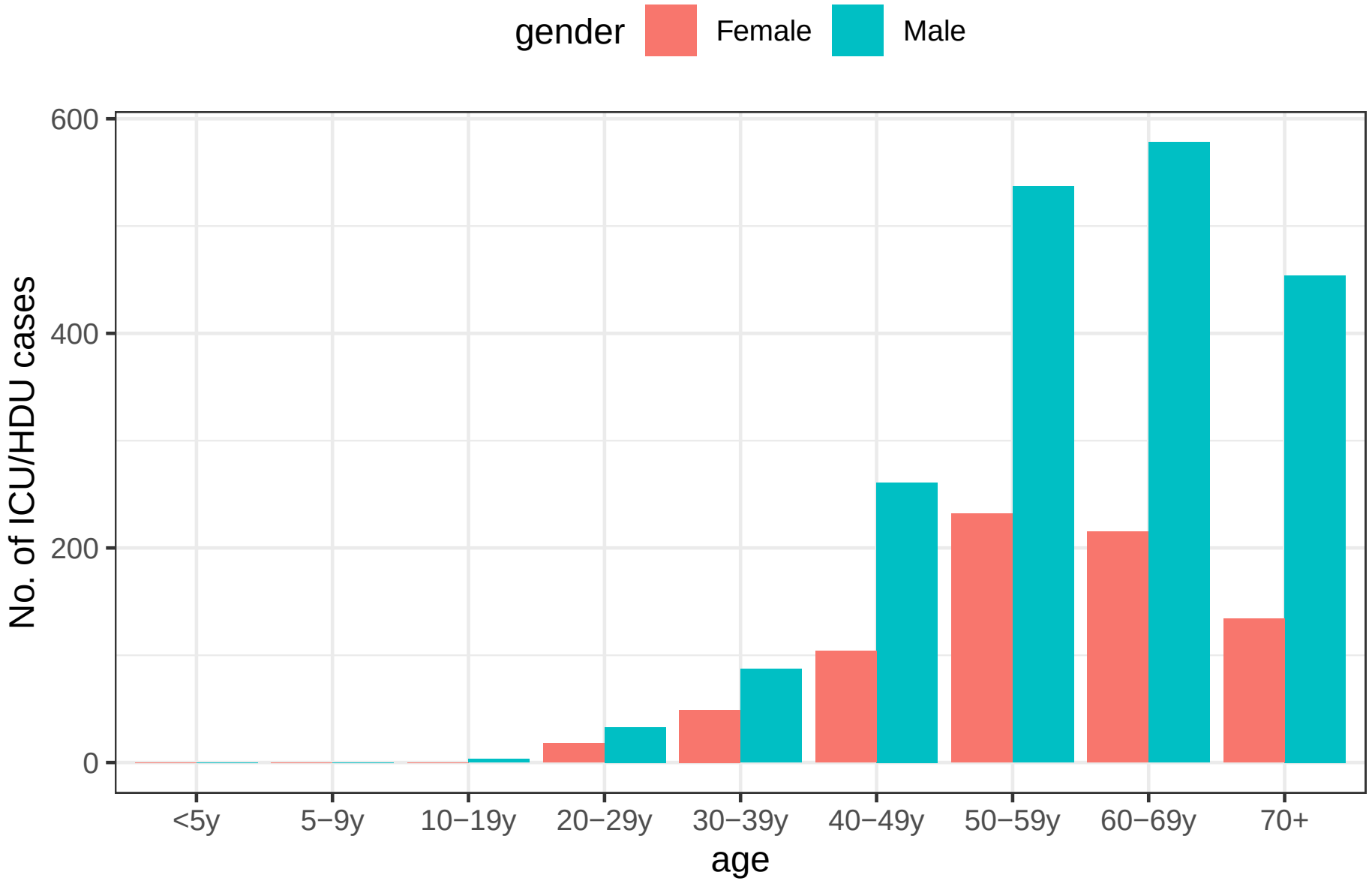
Hospitalised cases (L) and ICU/HDU cases (R).
PHE, data from England.





where men are much more
likely to have severe disease

Alternative: dodged barchart

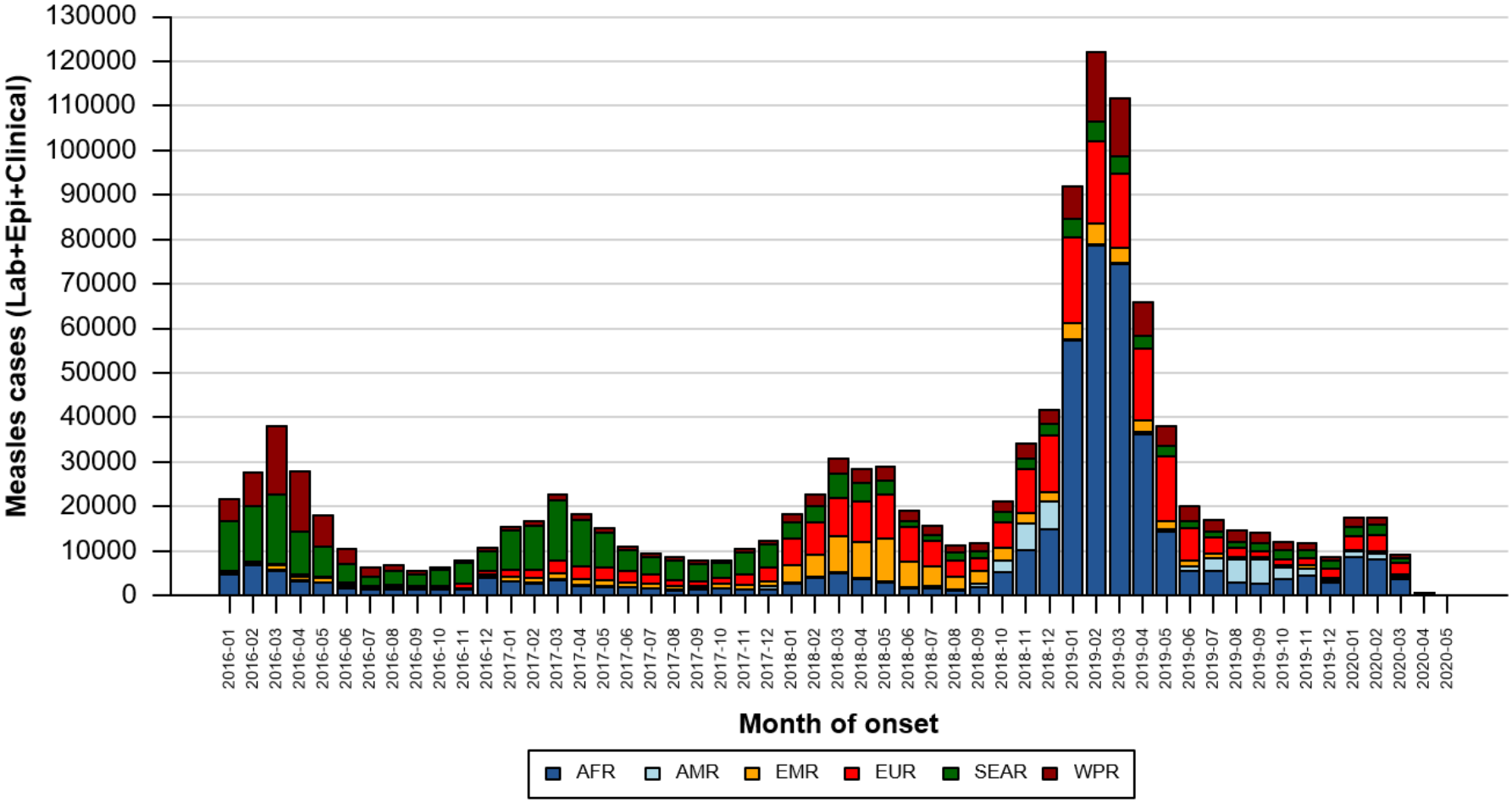


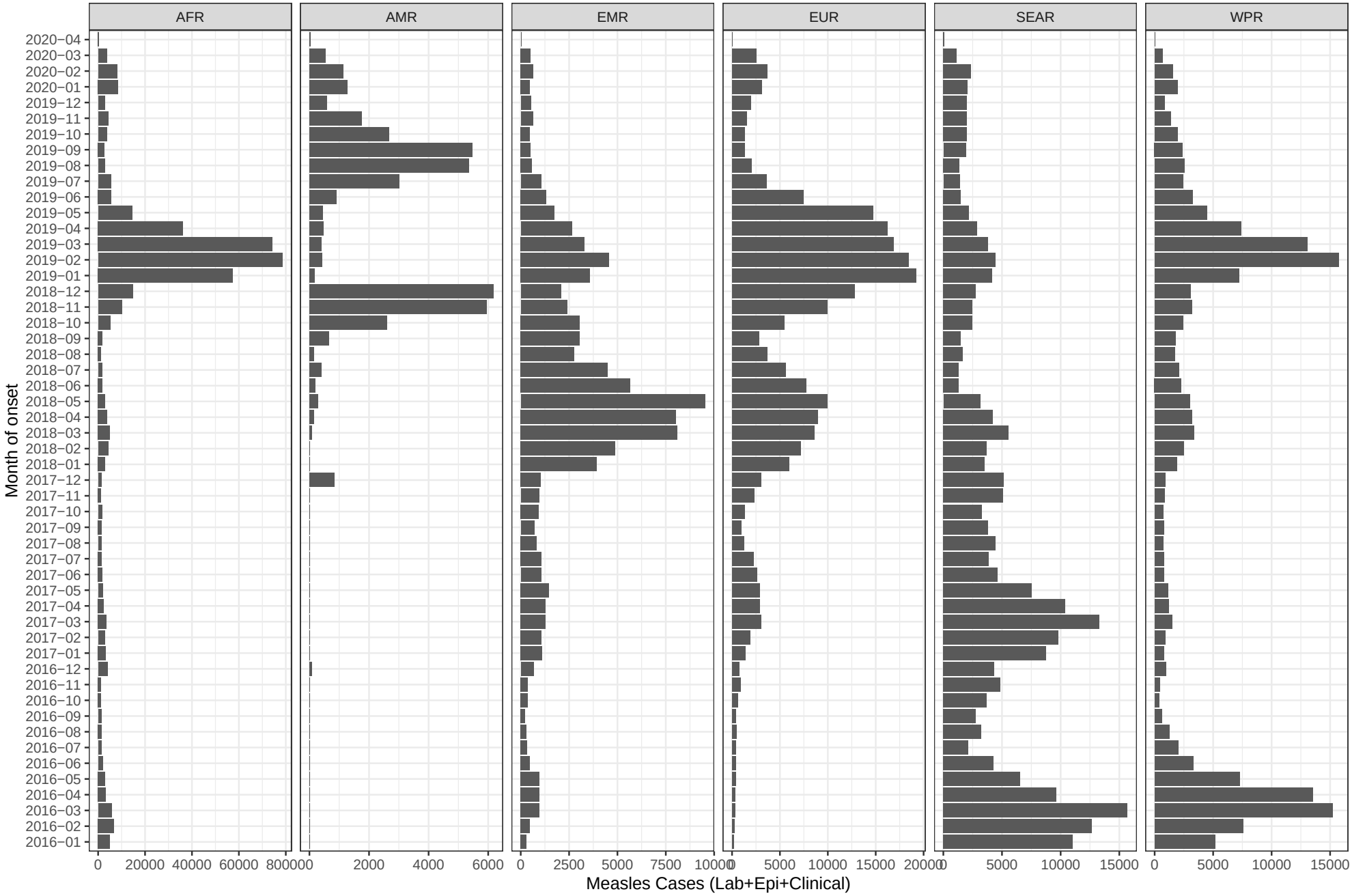
Decoding

- Table look-up: attach value to what is shown; pattern recognition
- Size (numeric)
  In decreasing level of efficiency:
  1. Position along common scale
  2. Position along identical, nonaligned scales
  3. Length; grid lines are useful
- Group (categorical)

3. Length

Measles case distribution by month and WHO Region (2016-2020)

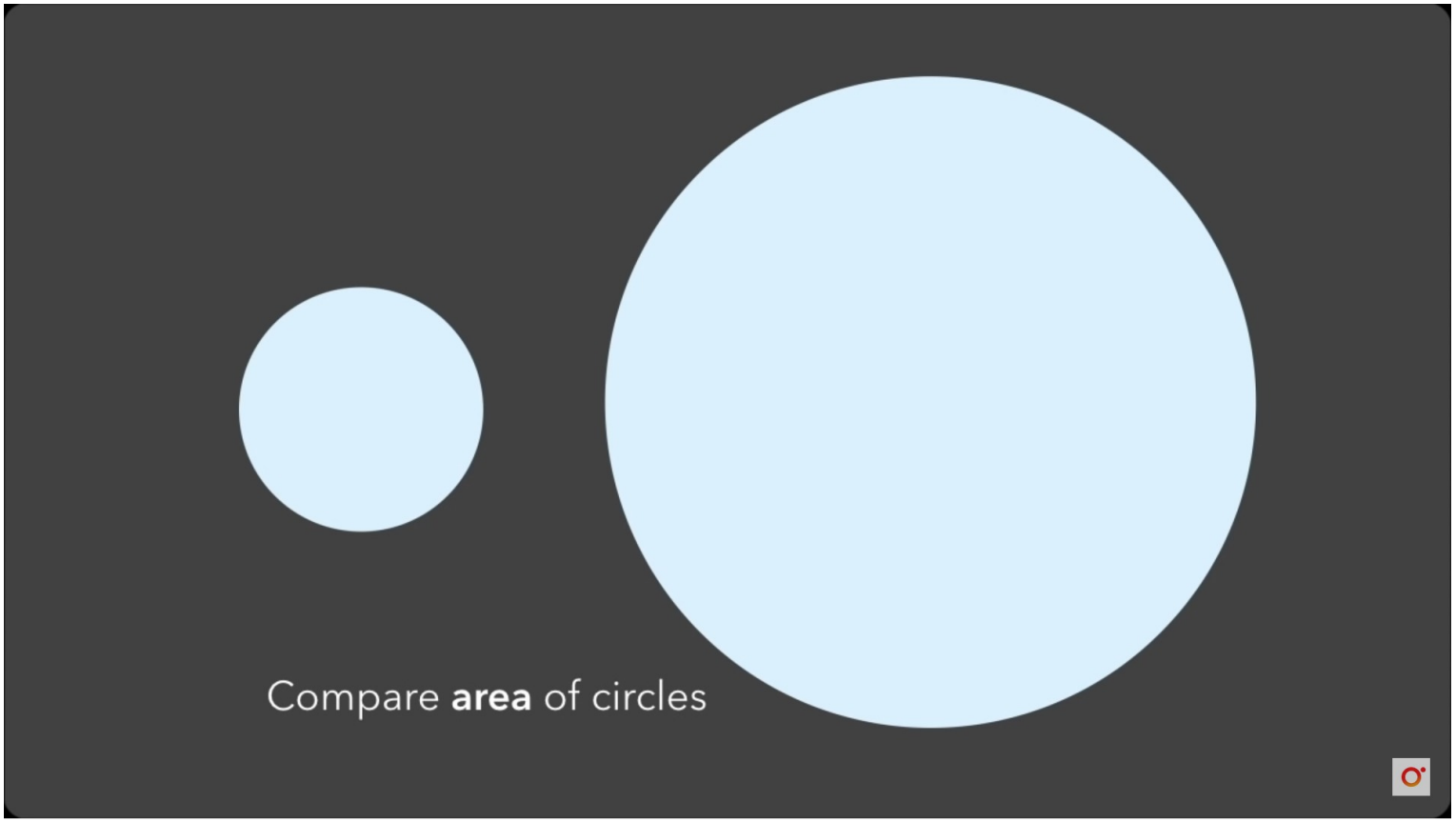



AFR
AMR
EMR
EUR
SEAR
WPR
Month of onset
2020−04
2020−03
2020−02
2020−01
2019−12
2019−11
2019−10
2019−09
2019−08
2019−07
2019−06
2019−05
2019−04
2019−03
2019−02
2019−01
2018−12
2018−11
2018−10
2018−09
2018−08
2018−07
2018−06
2018−05
2018−04
2018−03
2018−02
2018−01
2017−12
2017−11
2017−10
2017−09
2017−08
2017−07
2017−06
2017−05
2017−04
2017−03
2017−02
2017−01
2016−12
2016−11
2016−10
2016−09
2016−08
2016−07
2016−06
2016−05
2016−04
2016−03
2016−02
2016−01
0 20000 40000 60000 80000
0 2000 4000 6000
0 2500 5000 7500 10000
0 5000 10000 15000 20000
0 5000 10000 15000
0 5000 10000 15000
Measles Cases (Lab+Epi+Clinical)

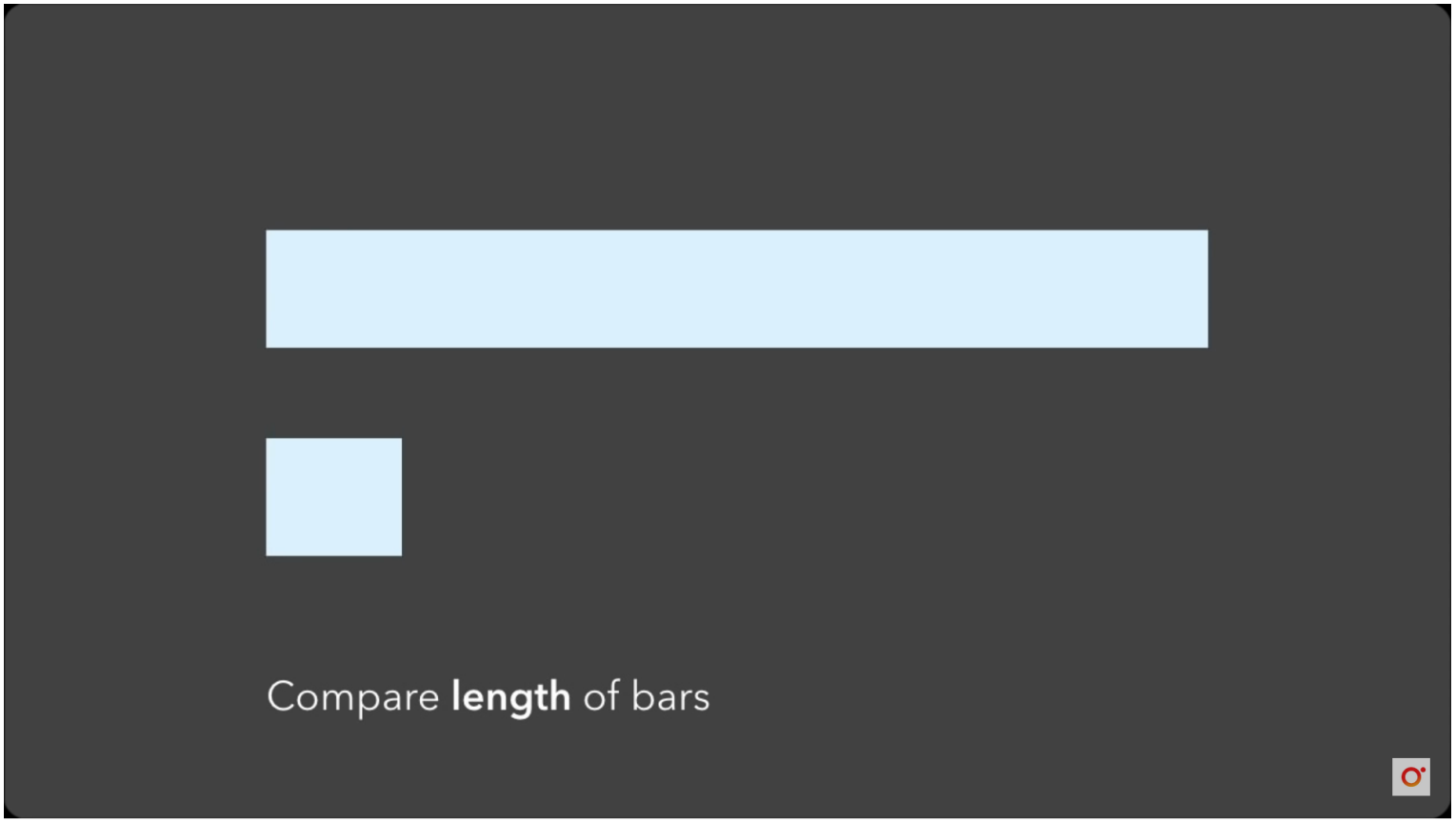
Decoding

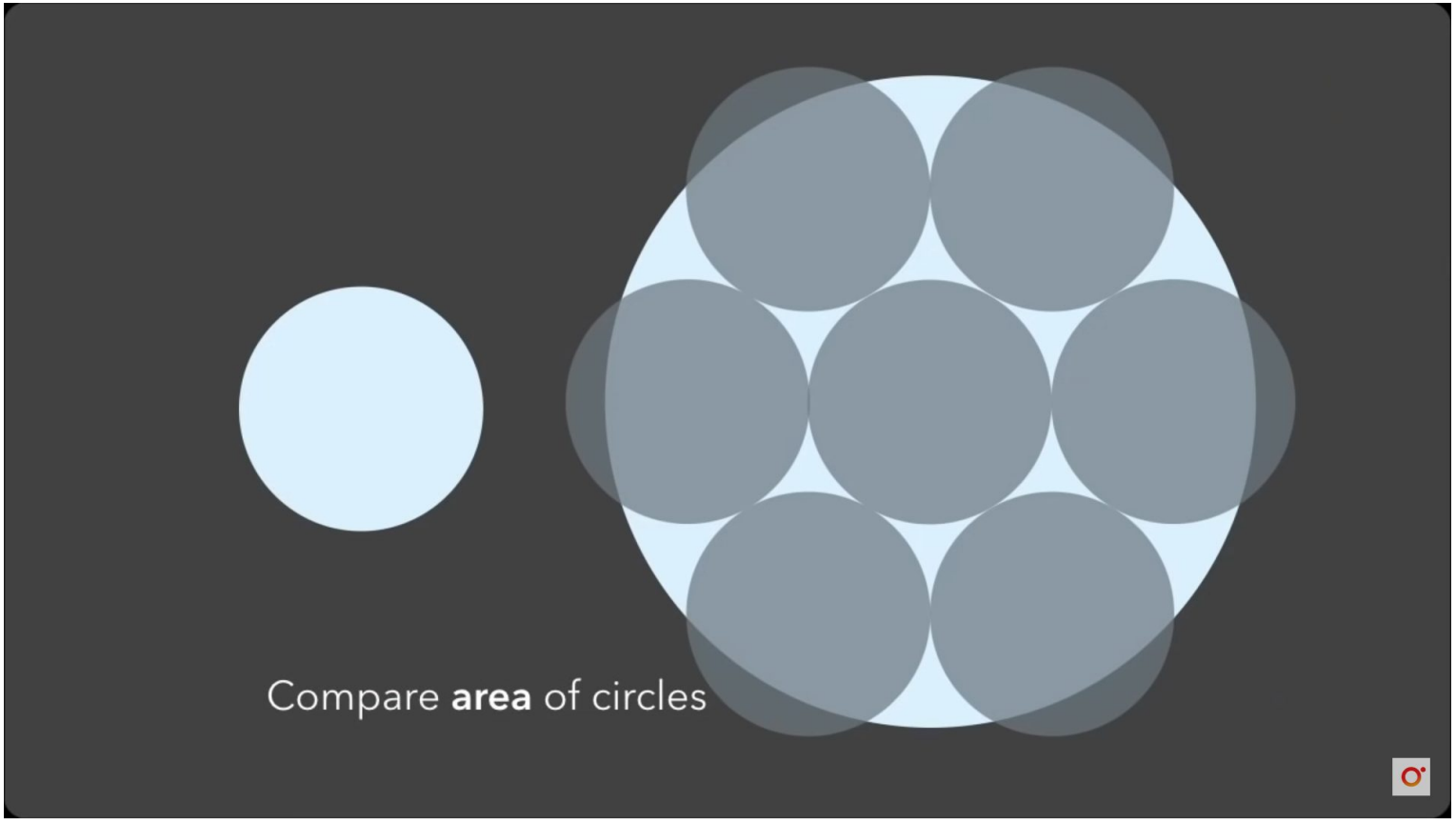
- Table look-up: attach value to what is shown; pattern recognition
- Size (numeric)
  In decreasing level of efficiency:
  1. Position along common scale
  2. Position along identical, nonaligned scales
  3. Length; grid lines are useful
  4. Angle, slope; don't use pie charts
  5. Area
- Group (categorical)

How many times larger is the second circle?

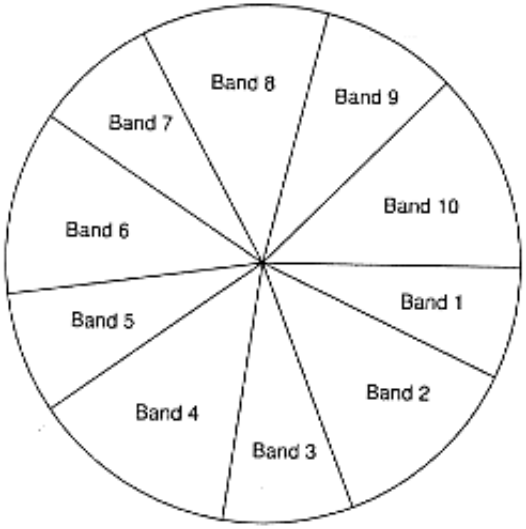


How many times larger is the first bar?

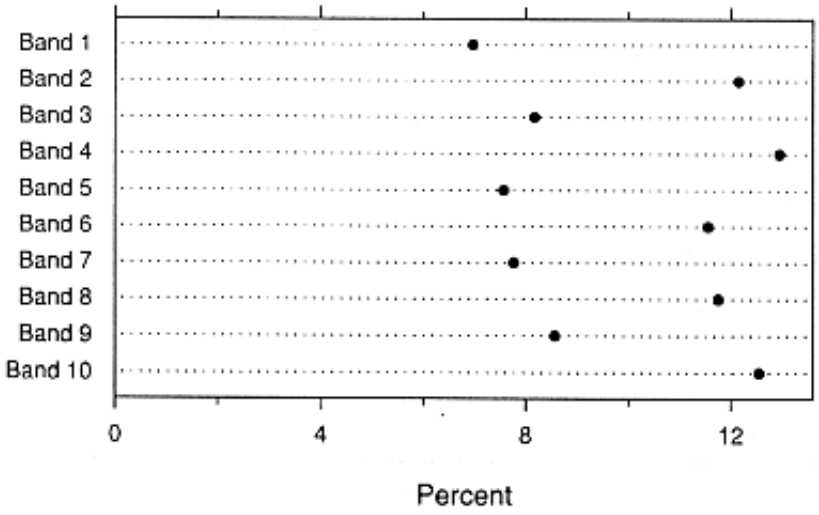



Compare **area** of circles

Angle: pie chart versus dot plot



4.19 PIE CHART. The pie chart falls in the category of a pop chart — a graphical method used frequently in the mass media and certain business presentations but far less in science and technology. Both table look-up and pattern perception are less efficient for pie charts than for dot plots.



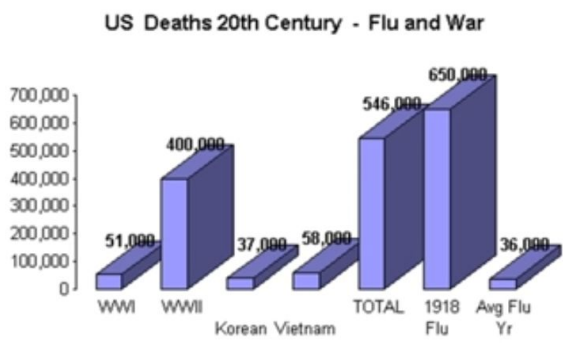
4.20 DOT PLOT. The data from Figure 4.19 are graphed by a dot plot. Patterns emerge that cannot be decoded from Figure 4.19.

Decoding

- Table look-up: attach value to what is shown; pattern recognition
- Size (numeric)
  In decreasing level of efficiency:
  1. Position along common scale
  2. Position along identical, nonaligned scales
  3. Length; grid lines are useful
  4. Angle, slope; don't use pie charts
  5. Area
  6. Volume; don't use unneccesary third dimension
- Group (categorical)

Mortality determined by fatality rate and how many infected.

- Infection fatality rate of COVID-19 probably around or just below 1% (depending on age structure of country- and how many asymptomatic).
- Comparison with other diseases:
- Ebola around 70% when first emerged.
- HIV 100% when first emerged.
- Smallpox 30%.
- H1N1 2009 flu 0.1%.
- H1N1 1918-20 'Spanish' flu around 3%.



if you combine all the people for example in the U.S.A.

GRESHAM COLLEGE

Decoding

- Table look-up: attach value to what is shown; pattern recognition
- Size (numeric)
  In decreasing level of efficiency:
  1. Position along common scale
  2. Position along identical, nonaligned scales
  3. Length; grid lines are useful
  4. Angle, slope; don't use pie charts
  5. Area
  6. Volume; don't use unneccesary third dimension
  7. Colour
- Group (categorical)

Alternative as heatmap

Chart 3: Daily **New** Coronavirus Cases Worldwide*

Country	12-Mar	13-Mar	14-Mar	15-Mar	16-Mar	17-Mar	18-Mar	19-Mar	20-Mar	21-Mar	22-Mar	23-Mar	24-Mar	25-Mar	26-Mar	27-Mar	28-Mar	29-Mar	30-Mar	31-Mar	1-Apr	2-Apr	3-Apr	4-Apr	5-Apr	6-Apr	7-Apr	8-Apr	9-Apr	10-Apr	11-Apr	12-Apr	13-Apr	14-Apr	15-Apr	16-Apr	17-Apr	18-Apr	19-Apr
Saudi Arabia	9	18	21	21	18	21	17	43	57	55	85	72	106	127	125	136	109	100	110	115	129	147	147	154	171	180	189	188	221	262	310	383	412	430	457	480	552	726	875
Morocco	1	1	4	6	6	8	8	9	12	15	17	20	23	32	40	51	58	64	70	68	63	57	59	76	92	103	98	89	88	82	90	97	97	110	[illegible]	[illegible]	[illegible]	[illegible]	208
Bangladesh	-	-	-	1	1	2	3	3	3	4	3	4	5	4	4	4	2	2	1	1	2	2	3	5	9	17	26	37	61	75	80	101	118	147	[illegible]	[illegible]	[illegible]	[illegible]	306
Singapore	7	13	13	12	16	17	25	30	36	42	36	41	43	50	57	56	61	53	49	49	50	51	59	66	77	82	92	109	150	183	205	227	252	286	35[illegible]	[illegible]	[illegible]	[illegible]	722
Russia	3	7	10	11	16	17	22	34	41	48	55	60	61	88	118	150	192	219	249	325	378	504	578	599	653	699	837	985	1,186	1,394	1,522	1,775	2,049	2,296	2,72[illegible]	[illegible]	[illegible]	[illegible]	4,591
Ukraine	-	1	1	1	2	3	3	3	6	8	15	14	17	25	31	59	65	83	88	84	110	106	131	145	129	106	98	111	146	221	262	277	303	292	313	346	390	434	421
Brazil	8	32	30	31	37	43	55	115	148	175	294	326	364	383	360	373	414	426	399	575	733	947	1,119	1,161	1,074	1,029	1,245	1,453	1,741	1,869	1,673	1,506	1,335	1,406	1,898	2,058	2,563	2,849	2,584
Pakistan	4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	91	124	119	105	118	83	106	125	131	134	129	141	156	206	242	220	260	336	337	361	333	232	244	242	252	286	343	422	382	450	491
Qatar	62	76	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8	11	11	10	14	14	14	15	16	24	36	55	61	79	96	136	192	221	246	221	193	170	168	192	214	229	246	281	358	395	434
Belarus	2	5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8	10	6	8	3	3	3	3	3	2	17	15	17	53	50	72	100	99	128	157	231	320	341	378	358	325	376	407	465	375	263
Peru	2	5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	37	50	55	40	46	41	54	60	64	93	93	108	163	141	161	170	240	287	340	649	744	834	974	794	1,132	1,102	1,157	1,243	926	1,029	1,038
India	9	10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	31	47	60	76	73	82	83	97	113	92	131	128	253	380	329	421	398	559	686	709	784	705	784	822	932	972	969	1,056	975	1,059	1,323
Serbia	5	9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	20	27	35	37	42	53	41	52	89	89	100	111	100	108	173	181	212	257	243	261	240	226	233	241	297	340	373	422	409	382	361
Mexico	1	1	5	8	10	18	17	19	28	30	40	50	51	51	56	67	88	111	130	127	125	133	129	149	169	191	232	274	323	325	351	359	370	393	389	407	409	465	525
Japan	34	48	48	50	50	44	29	21	31	32	53	51	58	75	72	85	125	140	120	121	121	157	188	297	240	290	322	280	382	469	525	623	676	529	524	470	604	663	674
UAE	10	10	3	6	3	3	7	11	11	14	10	15	27	45	45	52	55	59	70	65	87	114	163	210	246	263	274	289	298	321	344	366	383	393	407	426	445	342	354
Ireland	6	17	24	22	32	33	41	107	129	141	154	142	162	195	228	249	272	263	273	279	258	309	341	342	387	379	359	368	395	681	805	895	1,018	848	905	904	833	820	676
Indonesia	7	13	17	21	25	26	33	49	59	70	72	67	79	85	95	117	117	124	130	121	131	126	143	141	149	175	188	216	255	255	276	321	316	332	324	319	342	352	360
Egypt	5	6	13	13	21	29	22	37	34	25	33	28	29	41	42	43	44	38	40	44	51	64	82	90	99	114	116	123	132	118	122	126	123	139	142	152	164	171	160
Dom. Rep	-	-	2	2	2	4	3	6	15	23	45	53	60	70	72	84	102	117	103	132	141	130	147	95	115	112	117	156	151	198	201	214	205	167	214	197	240	262	267
Turkey	0	1	1	1	4	11	23	47	85	156	285	334	378	441	598	1,042	1,383	1,696	1,800	1,958	2,069	2,230	2,524	2,601	2,848	3,021	3,297	3,573	3,803	4,203	4,515	4,683	4,692	4,521	4,306	4,309	4,374	4,305	4,229
Moldova	1	1	2	5	5	6	5	7	11	13	16	15	15	17	21	23	27	29	30	39	48	61	73	100	110	115	116	106	106	118	126	122	106	124	122	123	138	111	106
UK	46	120	190	172	273	290	375	393	616	777	776	1,003	1,038	1,143	1,517	2,005	2,287	2,535	2,660	2,684	3,138	3,598	4,059	4,249	4,643	4,527	4,315	4,749	4,359	5,582	5,981	5,933	5,925	5,060	4,902	4,735	5,050	5,117	5,422
Romania	9	19	25	22	27	24	34	37	38	46	43	75	122	135	149	179	165	227	270	238	252	231	269	342	351	330	309	287	335	353	393	385	358	353	307	352	359	385	383
Argentina	2	5	4	7	9	9	11	13	18	23	47	51	65	57	59	72	76	90	80	116	91	97	111	99	99	105	91	66	86	105	87	107	103	76	117	107	115	120	99
Poland	10	13	20	22	32	43	37	59	62	75	96	99	119	129	147	160	184	203	209	231	229	271	332	329	387	367	366	395	368	386	377	367	340	312	307	311	361	385	426
US	286	399	442	555	742	1,061	1,264	2,562	3,660	4,795	6,373	7,524	8,616	10,045	12,640	14,454	16,932	18,783	19,499	21,629	22,977	25,713	28,439	30,170	30,925	30,726	30,480	30,050	31,427	32,467	32,223	31,565	29,460	27,784	27,489	28,070	29,772	31,132	30,684
Canada	13	29	30	36	75	71	115	137	132	200	203	322	462	494	643	649	697	757	839	961	996	1,251	1,260	1,113	1,549	1,320	1,359	1,541	1,225	1,374	1,361	1,289	1,256	1,244	1,223	1,628	1,784	1,830	1,856
Sweden	99	142	152	131	126	94	80	104	134	143	164	152	162	191	227	256	290	294	297	342	375	467	526	502	471	410	391	494	578	620	615	516	452	440	444	514	567	594	615
Chile	4	9	12	13	33	40	44	41	70	84	99	127	122	151	169	216	247	249	286	282	281	316	322	356	360	353	345	346	375	422	453	417	388	354	337	399	432	453	454
France	289	616	676	560	1,098	1,009	1,157	1,610	1,519	1,687	1,780	2,288	2,466	2,784	3,327	3,320	3,871	3,777	3,905	4,856	4,911	4,805	5,008	4,168	3,416	3,854	3,490	3,389	3,989	4,099	3,925	9,664	9,507	9,906	9,930	6,345	5,934	4,447	4,879
Hungary	2	3	5	5	7	8	7	10	12	13	18	24	26	31	33	33	39	46	47	48	46	44	44	47	52	40	49	54	62	112	123	129	120	81	67	61	76	81	84
Netherlands	60	121	145	159	228	226	274	332	397	482	540	574	644	700	813	971	1,060	1,123	1,087	1,005	969	965	1,001	1,015	1,064	1,035	972	989	988	1,081	1,216	1,266	1,202	1,083	936	909	977	1,047	1,131
Ecuador	1	1	3	3	5	10	21	43	83	112	170	196	179	167	154	154	185	188	140	161	231	310	352	306	225	146	95	246	330	854	878	754	641	111	150	190	230	355	403
Panama	3	7	9	9	11	11	13	17	21	33	57	59	52	61	61	82	110	115	108	127	99	104	122	123	155	168	156	144	182	191	219	246	218	180	150	129	154	185	175
Belgium	29	80	106	143	186	171	199	227	300	393	479	487	503	531	709	885	1,216	1,475	1,416	1,373	1,208	1,128	1,218	1,414	1,432	1,367	1,356	1,243	1,323	1,463	1,456	1,561	1,402	1,113	1,389	1,291	1,387	1,516	1,231
Colombia	2	3	5	6	11	13	18	17	19	33	35	44	63	69	65	66	58	58	77	92	114	115	117	125	105	105	128	162	185	224	232	181	157	127	99	114	147	115	172
Finland	9	31	46	46	55	42	28	39	43	51	73	75	86	89	83	85	94	90	99	94	70	70	66	116	120	165	173	151	170	148	149	122	115	98	83	99	106	130	137
South Africa	4	5	8	10	11	10	20	25	35	45	40	63	88	117	163	192	158	143	100	46	48	46	45	58	69	56	61	65	70	79	70	82	85	103	120	108	128	155	163
Algeria	1	2	4	7	8	9	9	10	9	20	32	36	44	41	42	45	48	52	54	77	98	119	147	134	118	109	74	80	87	85	89	86	79	77	84	89	109	116	117
Philippines	11	11	20	23	23	31	23	19	22	30	45	61	81	82	82	85	131	196	210	320	309	304	368	253	234	257	187	194	208	134	166	195	214	257	256	253	237	216	202
Denmark	145	178	143	108	79	55	70	88	101	99	100	87	95	111	127	157	162	176	183	210	231	252	298	308	318	326	330	332	317	285	231	193	171	173	171	176	189	183	176
Portugal	7	21	32	47	68	84	70	135	172	208	288	319	336	429	486	552	702	742	716	794	770	768	870	770	757	674	639	654	670	936	886	861	745	494	526	564	522	559	529
Israel	10	16	20	34	30	31	37	54	78	116	145	161	177	414	453	491	595	470	501	581	618	653	683	623	585	512	455	388	385	376	374	435	405	410	440	403	349	305	248
Malaysia	13	20	27	70	104	119	138	118	116	128	129	155	149	153	181	161	174	169	149	151	147	162	177	179	189	169	158	159	142	138	142	141	147	160	136	125	109	80	79
Croatia	2	5	6	8	10	8	11	14	18	35	43	53	64	59	60	68	69	68	74	70	77	75	72	65	55	53	51	54	56	68	63	64	61	52	52	48	41	32	33
Iran	877	1,051	1,172	1,235	1,229	1,201	1,158	1,117	1,163	1,110	1,069	1,161	1,292	1,602	1,942	2,321	2,649	2,823	3,022	3,068	3,046	3,040	2,922	2,785	2,658	2,508	2,352	2,211	1,999	1,923	1,860	1,775	1,771	1,671	1,590	1,577	1,548	1,498	1,456
Germany	260	625	782	972	1,299	1,396	1,936	2,381	3,144	3,239	3,137	3,434	3,285	3,778	4,766	5,454	6,177	6,193	5,737	5,234	5,044	5,675	6,069	6,071	5,563	4,645	4,126	4,301	4,515	4,699	4,311	3,640	2,973	2,297	2,461	2,461	2,831	2,996	2,608
New Zealand	-	-	0	1	1	2	4	5	8	10	21	19	29	38	45	67	74	77	77	70	64	71	70	76	83	77	73	65	50	44	38	30	28	21	19	18	15	14	11
Czechia	16	28	37	41	51	64	69	110	134	150	164	136	140	165	201	261	309	291	269	257	219	260	273	291	270	241	232	210	246	228	204	170	123	95	96	111	123	124	133
Greece	7	29	35	58	58	49	48	22	41	36	52	69	62	73	67	68	80	84	80	87	89	97	100	90	80	53	55	53	55	64	62	58	48	40	28	23	20	16	11
Spain	401	1,040	1,174	1,380	1,916	1,629	1,880	2,541	2,617	3,407	3,715	4,293	4,869	6,035	7,255	7,646	8,338	7,649	7,543	7,551	7,721	7,989	7,811	7,561	6,882	6,153	5,686	5,513	5,394	5,400	5,271	4,653	4,219	3,567	3,654	4,529	5,185	4,796	5,258
Thailand	5	6	7	14	19	26	33	40	44	59	97	112	126	131	112	104	105	114	120	129	132	122	114	104	100	86	70	76	64	63	65	46	39	35	31	30	30	30	31
Australia	13	27	36	42	62	63	80	96	104	155	245	250	313	323	315	365	399	405	388	354	306	283	242	248	206	170	141	115	105	105	102	76	61	50	34	37	43	33	27
Austria	50	93	118	154	179	207	248	288	343	371	484	615	724	694	832	796	747	800	677	631	610	585	477	400	335	292	279	290	298	315	292	251	199	168	133	133	139	111	103
Norway	132	198	173	156	158	117	115	131	145	164	209	219	237	242	246	284	288	300	269	222	212	216	231	227	206	180	179	134	131	112	81	110	98	77	83	93	84	103	85
Iceland	13	19	22	22	19	22	24	40	57	63	80	65	60	66	59	76	79	71	71	61	64	75	70	71	67	61	56	50	41	28	26	21	16	[illegible]	[illegible]	10	[illegible]	[illegible]	11
Luxembourg	4	8	12	13	15	27	38	69	102	133	149	135	154	166	164	183	183	154	134	143	122	134	156	138	121	89	90	76	78	95	75	62	44	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	44
Switzerland	79	191	217	387	387	390	417	469	774	969	1,112	1,180	1,146	1,081	1,084	1,033	1,050	983	1,028	919	923	1,000	921	975	833	708	662	694	738	724	714	534	409	346	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	351
Italy	1,272	2,122	2,752	3,071	3,880	3,462	3,639	4,072	4,760	5,518	5,856	5,723	5,539	5,202	5,363	5,643	5,824	5,826	5,288	4,824	4,526	4,388	4,522	4,710	4,594	4,326	3,940	3,698	3,670	3,758	4,171	4,235	3,973	3,728	[illegible]	[illegible]	[illegible]	[illegible]	454
China	27	21	23	21	25	28	31	38	54	62	83	86	85	89	87	100	102	115	104	96	91	78	78	66	60	58	52	67	70	69	74	81	83	91	86	67	137	120	112
South Korea	139	125	143	102	92	85	82	101	104	120	137	99	96	85	70	93	110	112	105	114	102	98	100	93	88	77	67	57	47	42	37	32	29	29	28	25	25	22	18

Greener cells represent the days with least daily new cases for a country

Countries with very recent outbreaks

Countries down here have left the worst behind

Reddest cells are the biggest number of new daily cases for a country

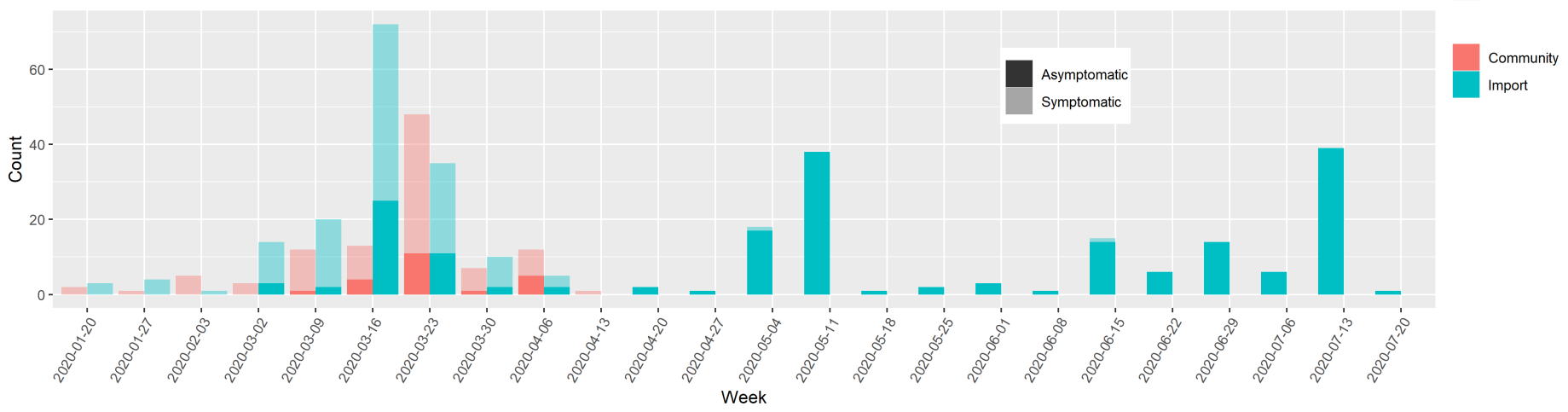
** As of April 19th 2020. Top ~60 countries in total cases, with added countries with special interest. Daily new cases calculated as 4-day averages to smooth outliers and weekend effects. Order based on average of last 5 days over average of last 25 days, adjusted manually to highlight the trend*
Source: Tomas Pueyo Analysis, Johns Hopkins data via Github: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv

oucru

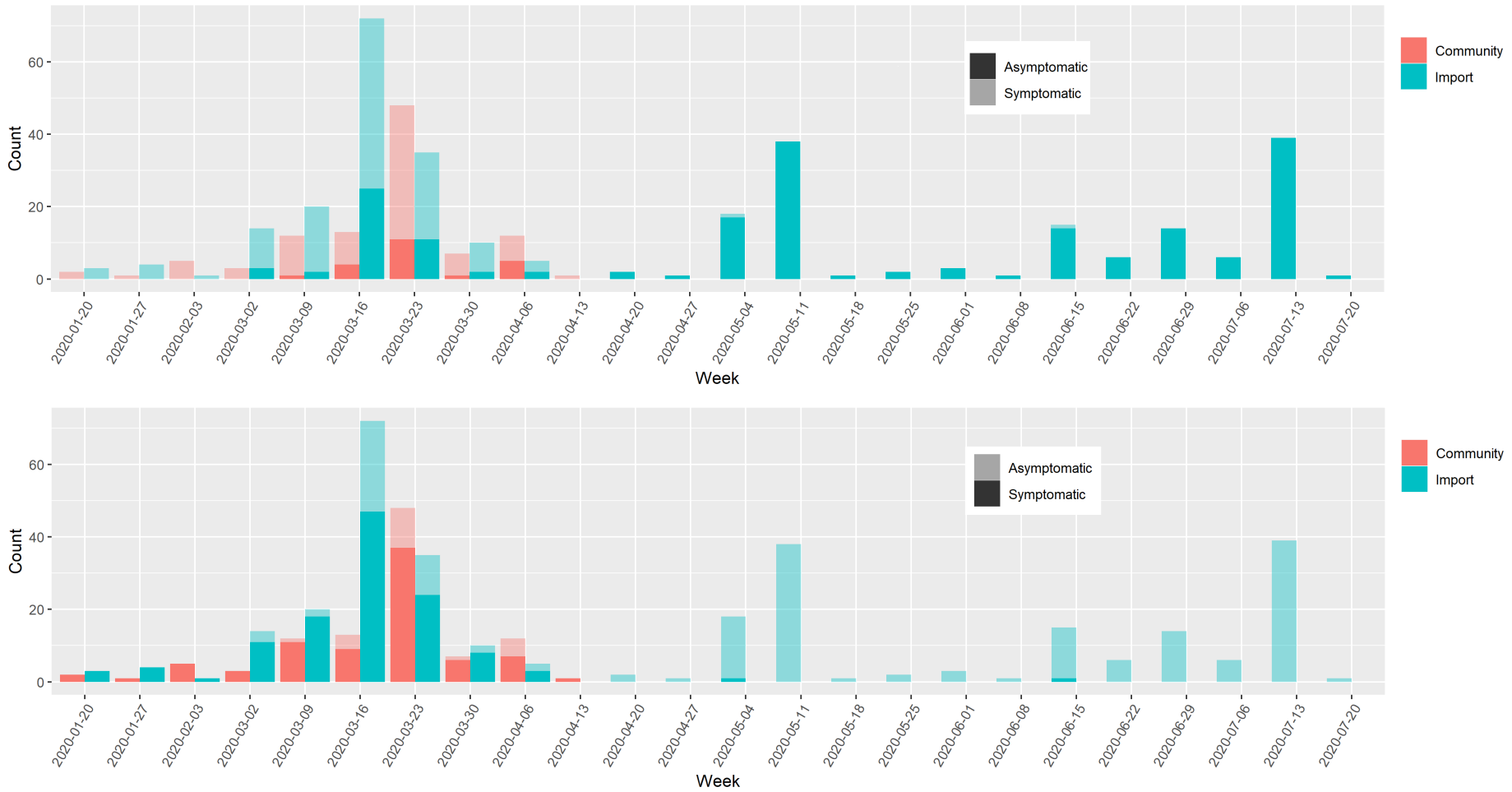
UNIVERSITY OF OXFORD

wellcome

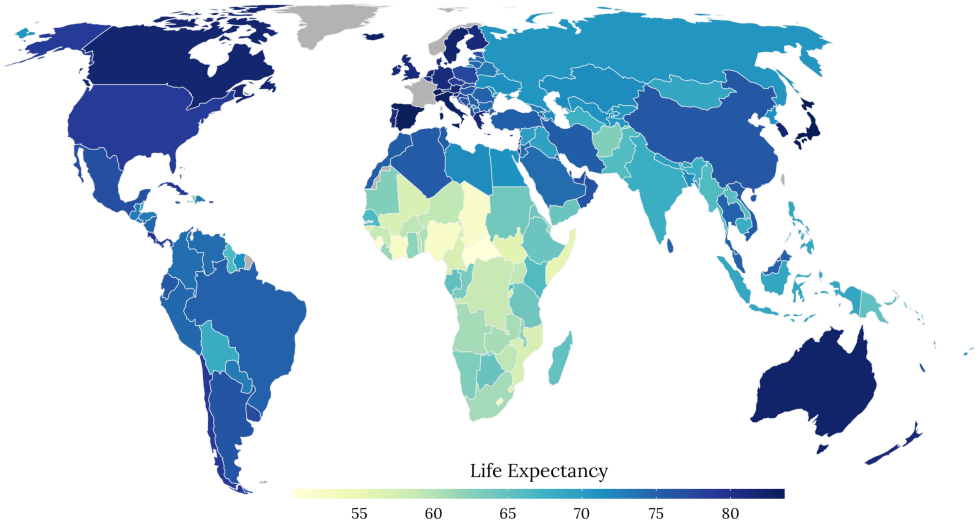
Number of SARS-CoV-2 infections in Vietnam

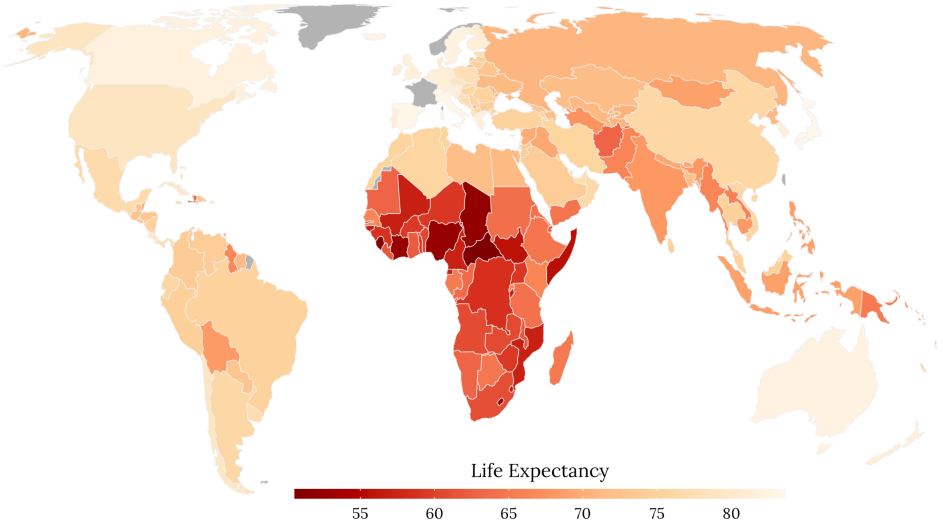


Number of SARS-CoV-2 infections in Vietnam



Which do you prefer?





Decoding

- Table look-up: attach value to what is shown; pattern recognition
- Size (numeric)
  In decreasing level of efficiency:
  1. Position along common scale
  2. Position along identical, nonaligned scales
  3. Length; grid lines are useful
  4. Angle, slope; don't use pie charts
  5. Area
  6. Volume; don't use unneccesary third dimension
  7. Colour
- Group (categorical)
  - Colour
  - Characters/symbols; "+" and "o" preferred for two groups
  - Linetypes (e.g. one thin black and one thicker gray line)

Outline

Introduction

Principles of Data Visualisation
- Scale and range
- Theories of perception

Decoding: William Cleveland
- A hierarchy of visual decoding
  - Size (numeric)
  - Group (categorical)

Colour

Aesthetics: Tufte

Final words

Basics

- Three colour models
  - RGB (Red-Green-Blue): how computers represent colour ($\approx$ the 3 eye cones in humans)
  - HSV (Hue-Saturation-Value): $\approx$ wavelength-colour purity-brightness
  - HCL (Hue-Chroma-Luminance): $\approx$ wavelength-colourfullness-brightness closest to perceptual properties

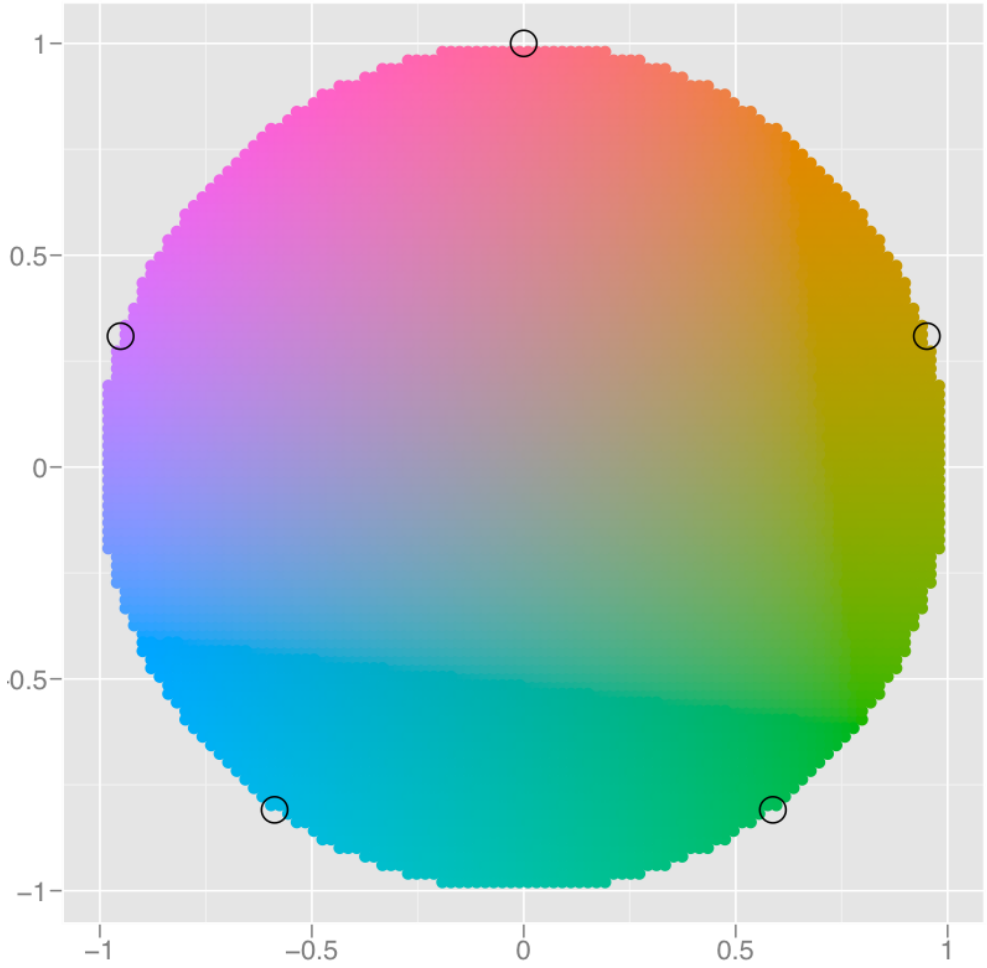
HCL explained

Hue
(Type of color)

Chroma
(Colorfulness)

Luminance
(Brightness)

HCL colour wheel (fixed luminance)



Basics

- Three colour models
  - RGB (Red-Green-Blue): how computers represent colour ($\approx$ the 3 eye cones in humans)
  - HSV (Hue-Saturation-Value): $\approx$ wavelength-colour purity-brightness
  - HCL (Hue-Chroma-Luminance): $\approx$ wavelength-colourfullness-brightness closest to perceptual properties
- Take colour-blindness into account: deuteranopia: green deficient; protanopia: red deficient; tritanopia: blue deficient

Basics

- Three colour models
  - RGB (Red-Green-Blue): how computers represent colour ($\approx$ the 3 eye cones in humans)
  - HSV (Hue-Saturation-Value): $\approx$ wavelength-colour purity-brightness
  - HCL (Hue-Chroma-Luminance): $\approx$ wavelength-colourfullness-brightness closest to perceptual properties
- Take colour-blindness into account: deuteranopia: green deficient; protanopia: red deficient; tritanopia: blue deficient
- In R colours are internally coded in hexadecimal format

```
     black       orange      skyblue bluishgreen
 "#000000"   "#E69F00"   "#56B4E9"   "#009E73"
```

R has 657 built in color names (see e.g. CheatSheet)

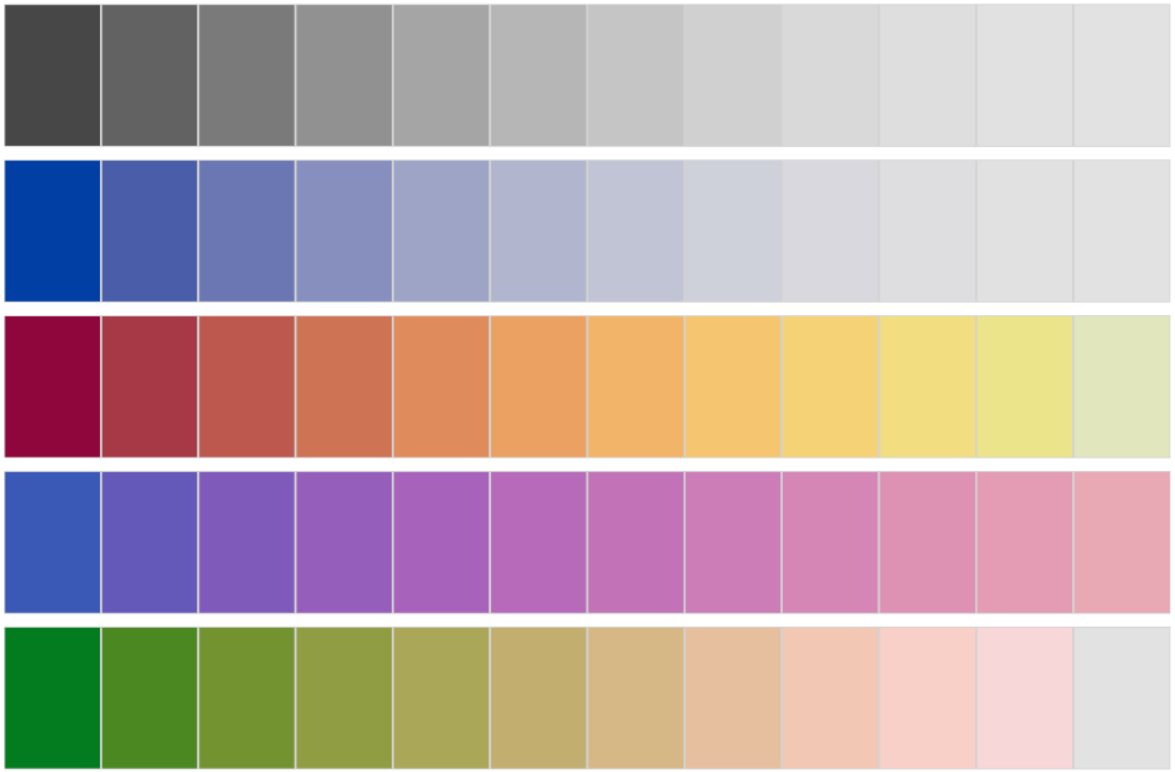
Represent value by colour: palettes

- Qualitative: categorical information
  - same perceptual weight to each category: vary hue, keep chroma and luminance fixed
  - may choose hues based on convention (red and blue for political parties in United States, cold/hot, risky/safe)

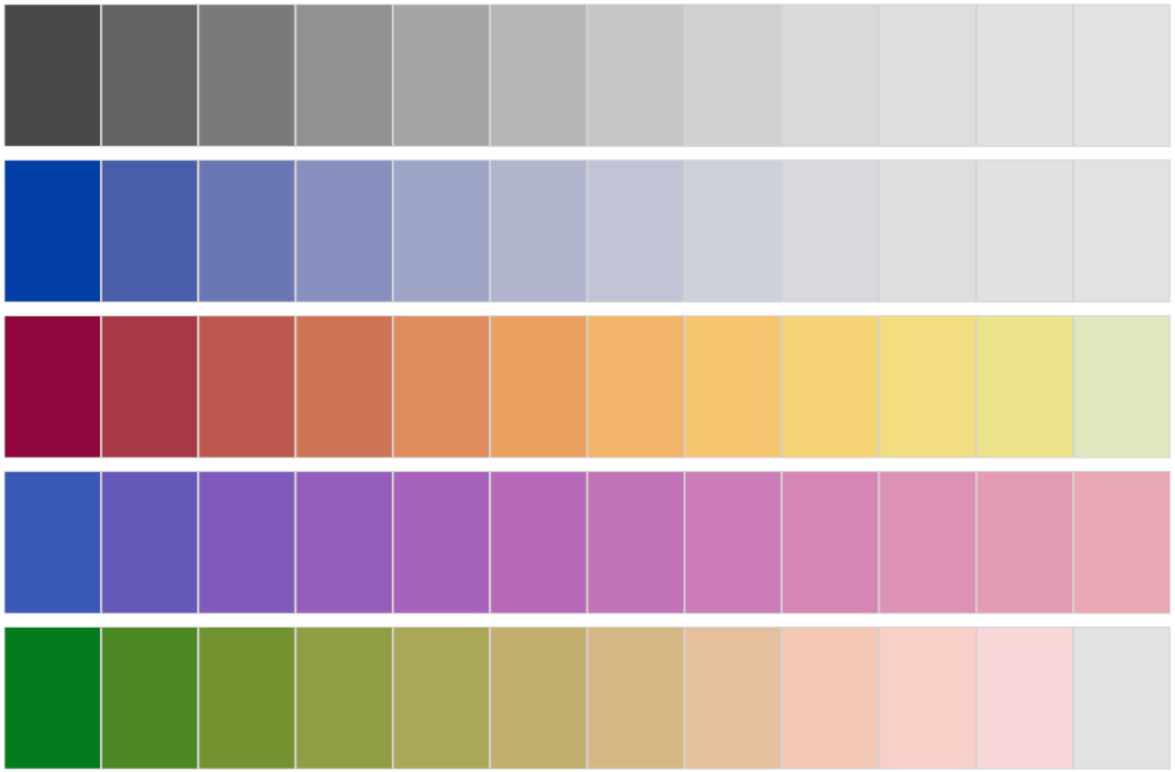
Represent value by colour: palettes

- Qualitative: categorical information
  - same perceptual weight to each category: vary hue, keep chroma and luminance fixed
  - may choose hues based on convention (red and blue for political parties in United States, cold/hot, risky/safe)
- Sequential: for ordered/numeric information
  - Single or two hues; change luminance linearly along with value, may change chroma

Sequential palettes

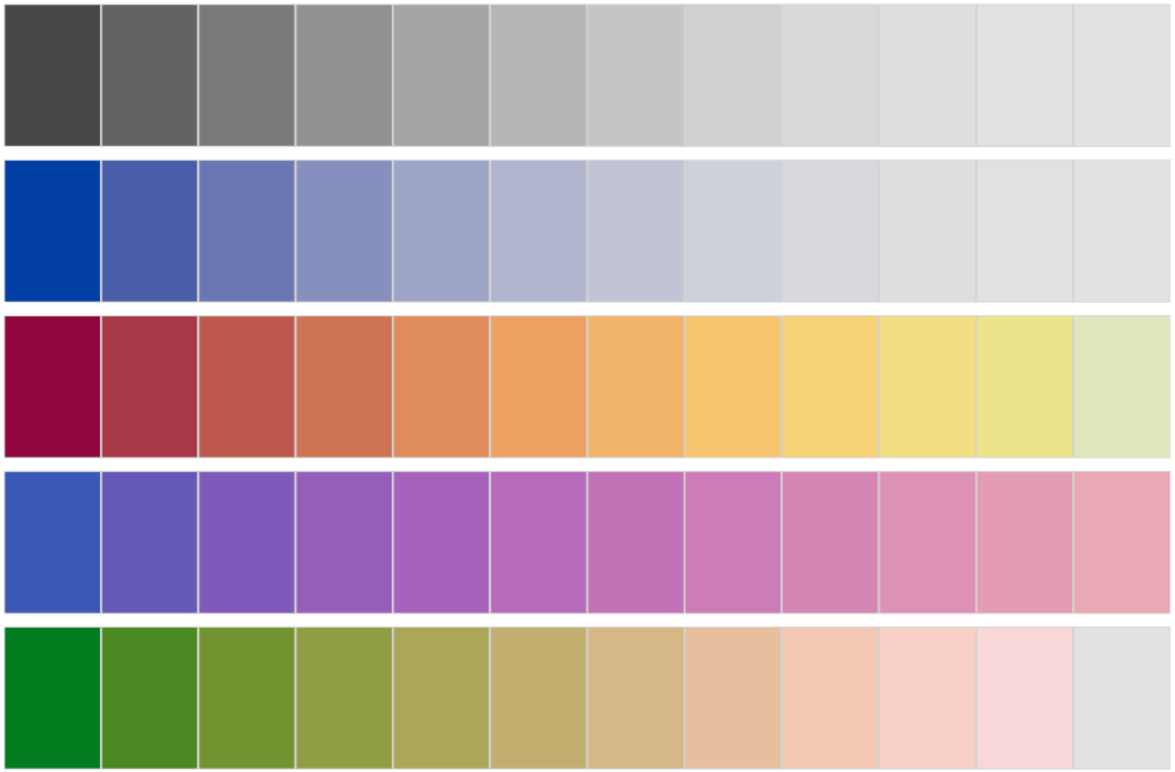


Sequential palettes



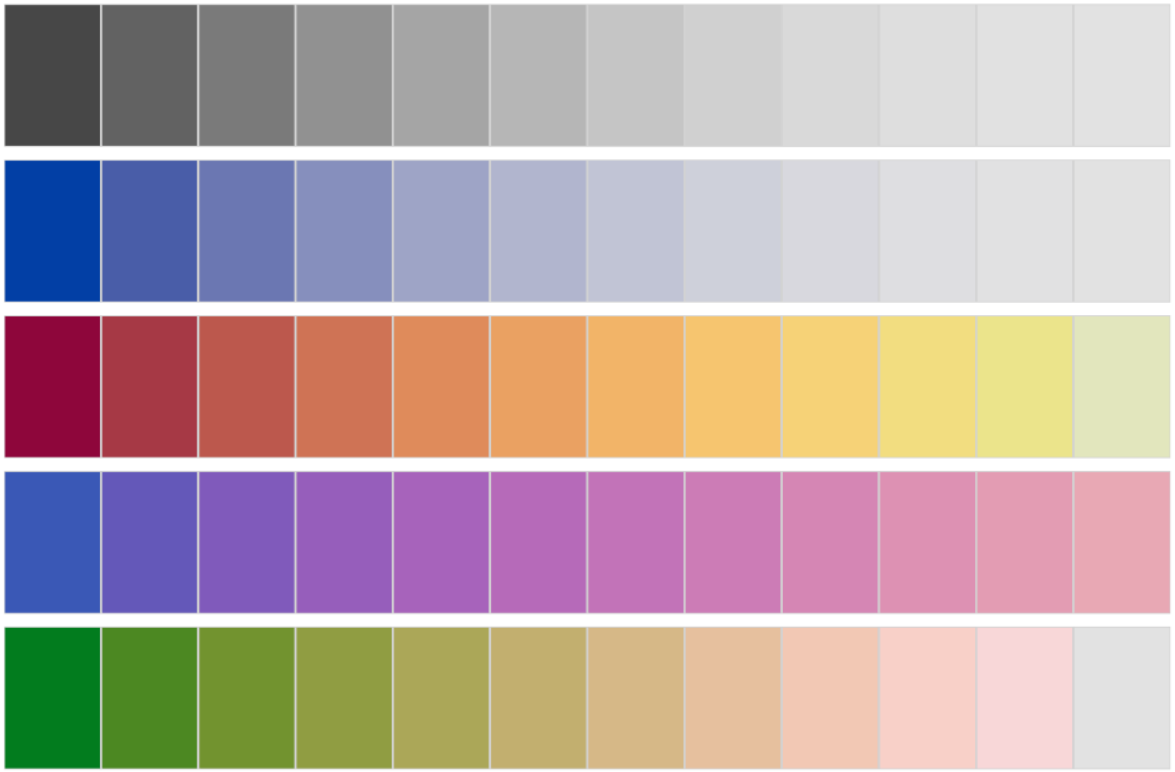
1st: vary luminance only

Sequential palettes



1st: vary luminance only; 2: vary chroma and luminance

Sequential palettes

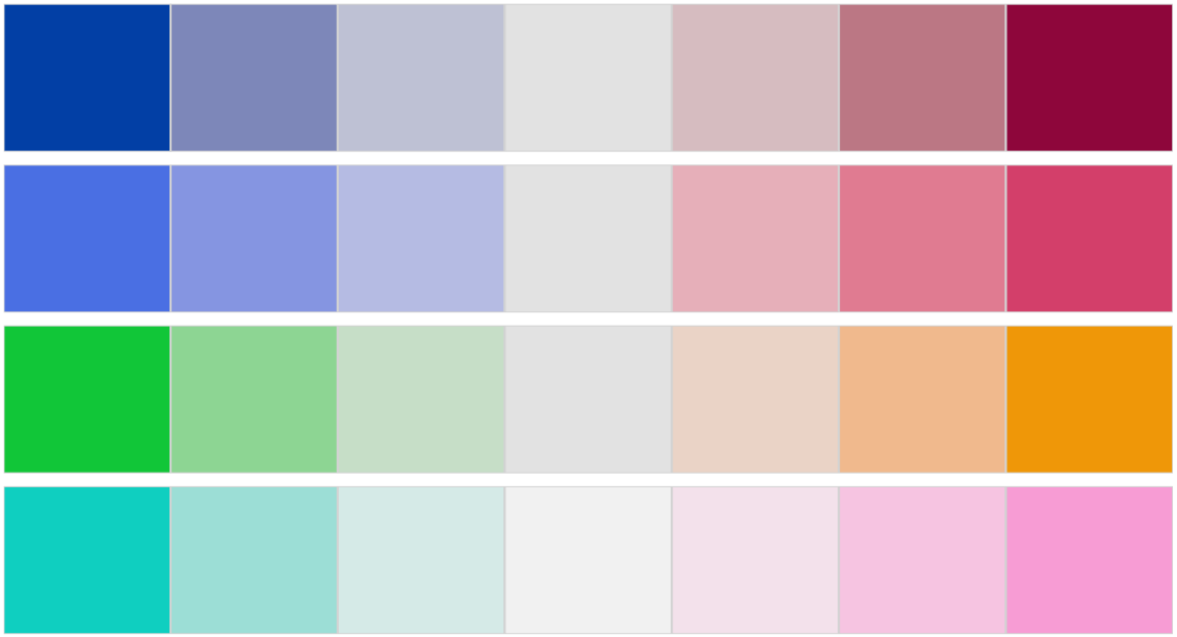


1st: vary luminance only; 2: vary chroma and luminance;
3-5: vary hue, chroma and luminance

Represent value by colour: palettes

- Qualitative: categorical information
  - same perceptual weight to each category: vary hue, keep chroma and luminance fixed
  - may choose hues based on convention (red and blue for political parties in United States, cold/hot, risky/safe)
- Sequential: for ordered/numeric information
  - Single or two hues; change luminance linearly along with value, may change chroma
- Diverging: for ordered/numeric information in two directions around neutral value
  Combines two sequential palettes with different hues

Diverging palettes



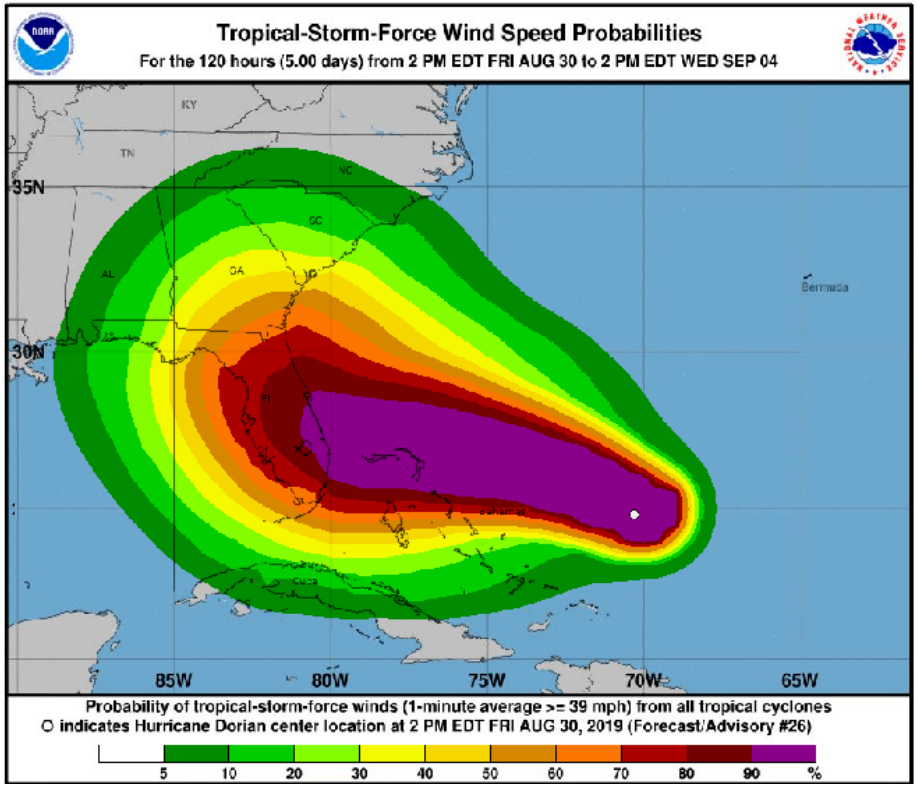
`hcl_palettes()` in `colorspace` package

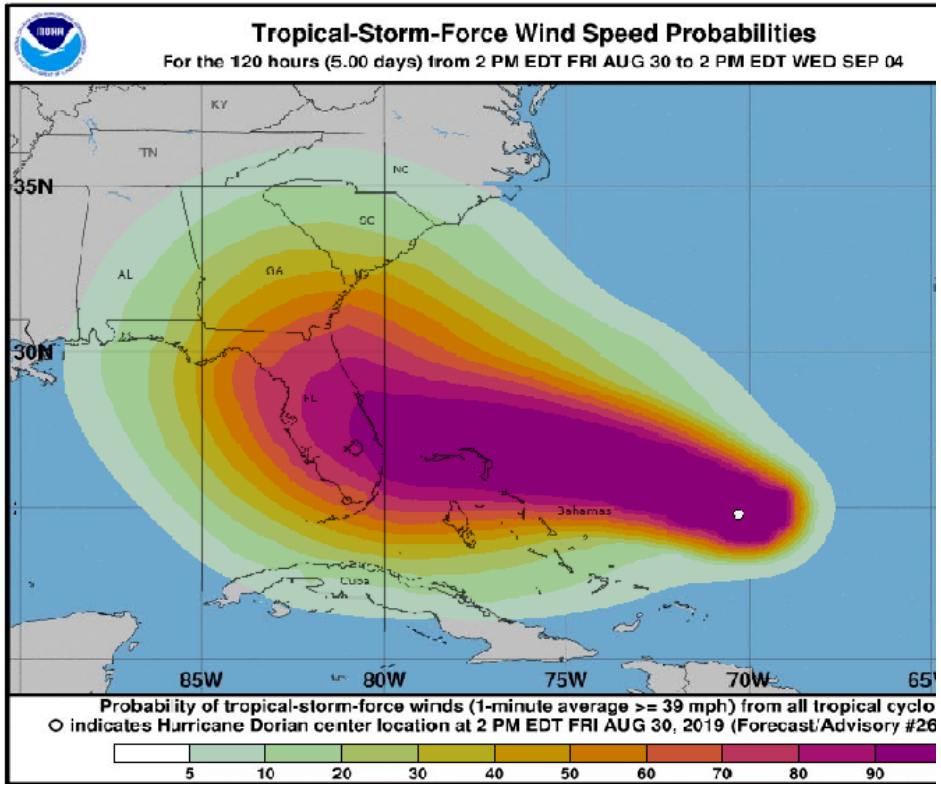
```
R> hcl_palettes(plot = TRUE)
```

Some advice on colour palettes

- Avoid large areas of flashy, highly saturated colors; use light colors (higher luminance, lower chroma). Recently changed in base R palette

Highly saturated colours may have confused Donald Trump





Some advice on colour palettes

- Avoid large areas of flashy, highly saturated colors; use light colors (higher luminance, lower chroma). Recently changed in base R palette
- For points and lines: lower luminance, higher chroma
- If size of coloured areas has a meaning, then use colours of similar luminance. Brighter colours tend to make areas look larger

Some advice on colour palettes

- Avoid large areas of flashy, highly saturated colors; use light colors (higher luminance, lower chroma). Recently changed in base R palette
- For points and lines: lower luminance, higher chroma
- If size of coloured areas has a meaning, then use colours of similar luminance. Brighter colours tend to make areas look larger
- `http://hclwizard.org/`: palette creator and colour deficiency emulator
- Top R colour palettes
- All R colour palettes (including interactive colour picker)
- Colour palette finder
- PrettyCols

Outline

Introduction

Principles of Data Visualisation
- Scale and range
- Theories of perception

Decoding: William Cleveland
- A hierarchy of visual decoding
  - Size (numeric)
  - Group (categorical)

Colour

Aesthetics: Tufte

Final words

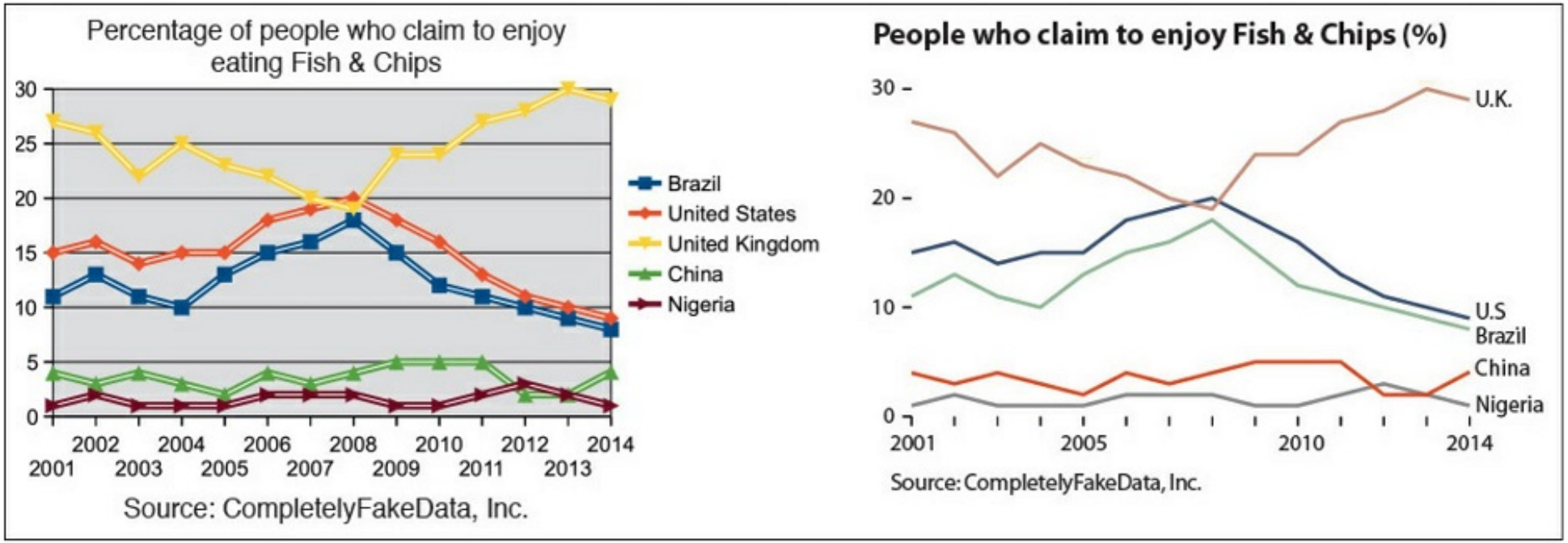
Criteria for graph quality

- Truthful
  - based on data collected in honest research
  - objective display of information
  - use appropriate scale and range
- Intuitive: easy to understand, concise without oversimplifying (see Section “Aesthetics: Tufte”)
- Informative (see Section “Decoding: William Cleveland”)
  - use appropriate graph type
  - use proximity and alignment to facilitate comparison
  - use labels and annotations to add clarity to the message
- Insightful: reveals evidence hard to observe otherwise (see Section “Aesthetics: Tufte”)
- Visually appealing; but honesty and clarity come first

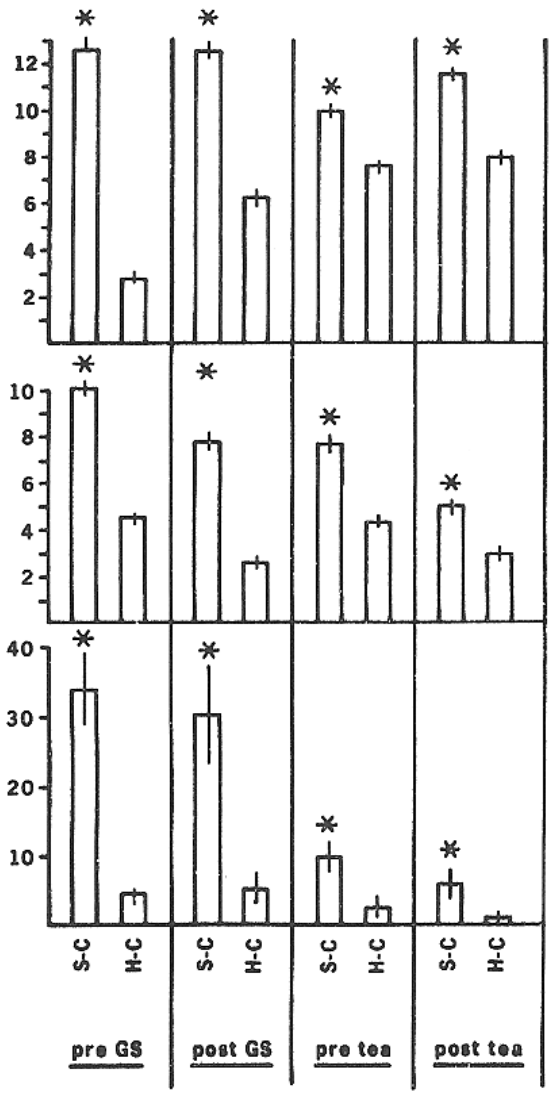
Edward Tufte: Graphical Excellence

- well-designed presentation of interesting data is a matter of substance, of statistics and of design.
- complex ideas communicated with clarity, precision, and efficiency
- is nearly always multivariate (table is better with < 30 data points)
- $$\text{data-ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}$$
  - dotplots preferred over barcharts
  - no unnecessary dimensions
- opposite of chartjunk

Which figure do you prefer?



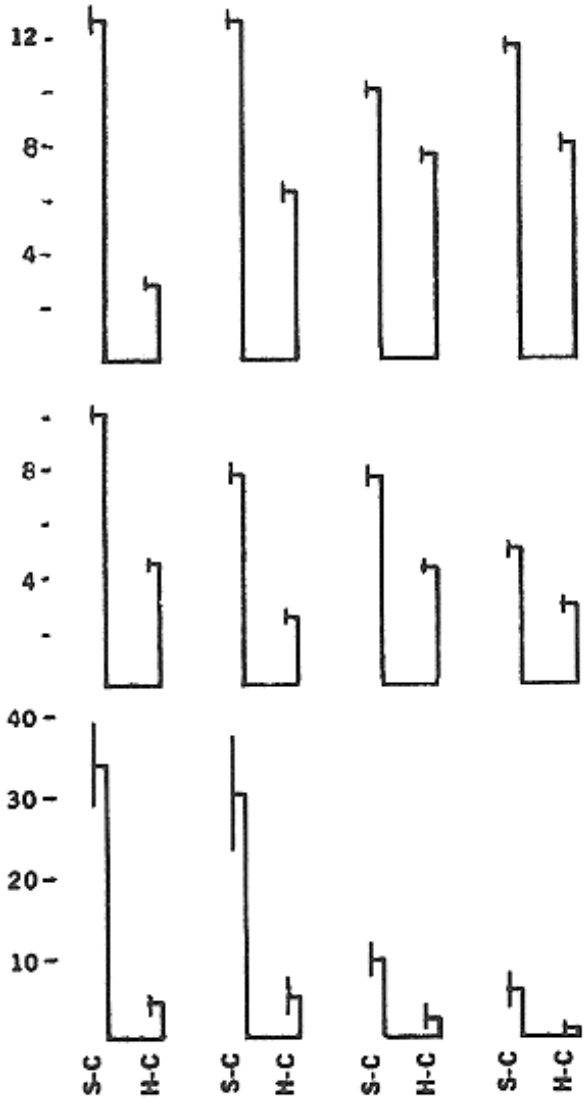
Example



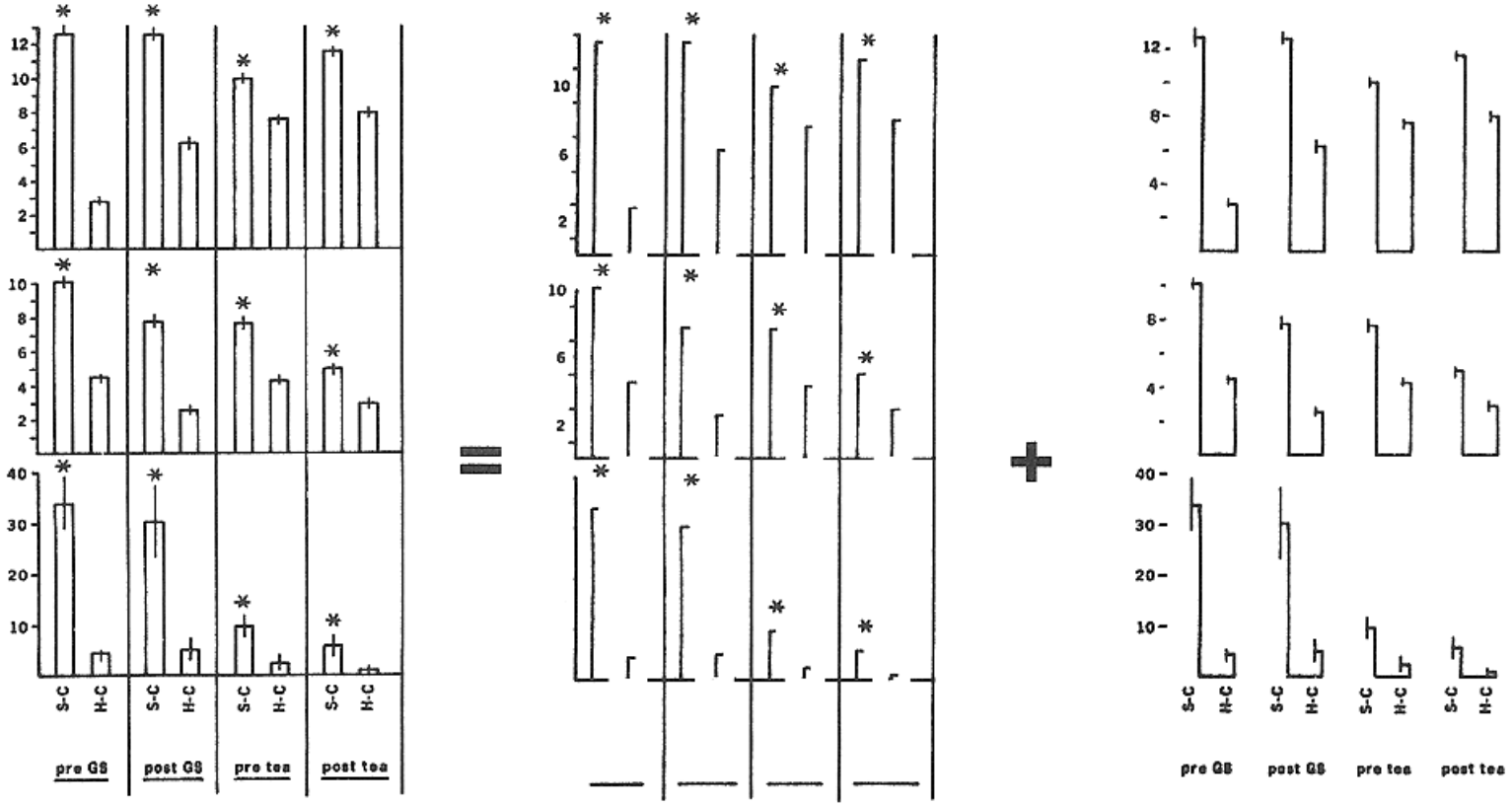
Cross-Enhancement of the Sour Taste on Single Human Taste Papillae

Geometric mean magnitude estimates ($\pm$ 1 s.e., 10 observations) of citric acid taste intensity (C) after sucrose (S) or water (H) pretreatment before and after treatment with Gymnema sylvestre (GS) or tea.

Noise removed

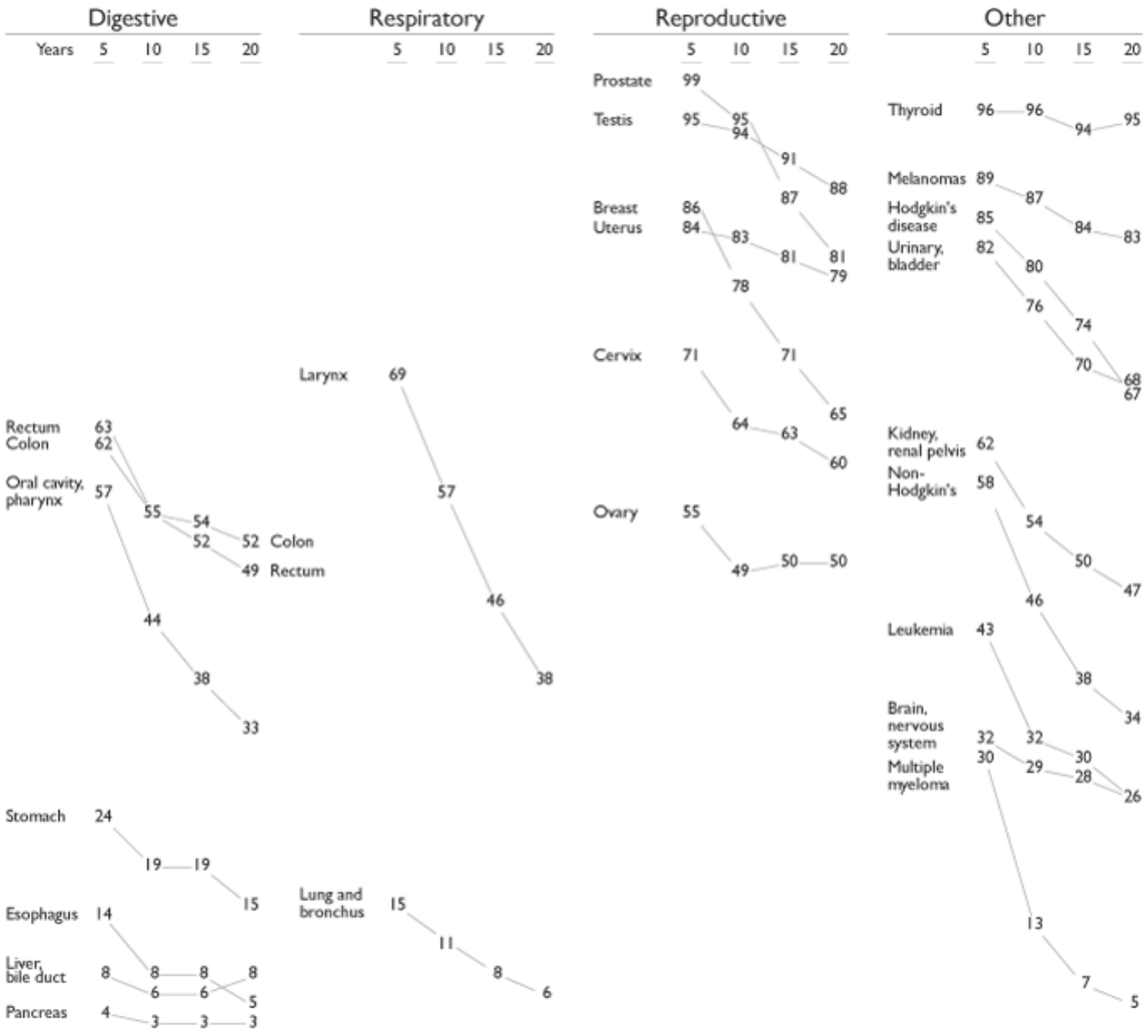


Hence



	Relative survival, % (SE)			
	5 years	10 years	15 years	20 years
Cancer site				
Oral cavity and pharynx	56.7 (1.3)	44.2 (1.4)	37.5 (1.6)	33.0 (1.8)
Oesophagus	14.2 (1.4)	7.9 (1.3)	7.7 (1.6)	5.4 (2.0)
Stomach	23.8 (1.3)	19.4 (1.4)	19.0 (1.7)	14.9 (1.9)
Colon	61.7 (0.8)	55.4 (1.0)	53.9 (1.2)	52.3 (1.6)
Rectum	62.6 (1.2)	55.2 (1.4)	51.8 (1.8)	49.2 (2.3)
Liver and intrahepatic bile duct	7.5 (1.1)	5.8 (1.2)	6.3 (1.5)	7.6 (2.0)
Pancreas	4.0 (0.5)	3.0 (0.5)	2.7 (0.6)	2.7 (0.8)
Larynx	68.8 (2.1)	56.7 (2.5)	45.8 (2.8)	37.8 (3.1)
Lung and bronchus	15.0 (0.4)	10.6 (0.4)	8.1 (0.4)	6.5 (0.4)
Melanomas	89.0 (0.8)	86.7 (1.1)	83.5 (1.5)	82.8 (1.9)
Breast	86.4 (0.4)	78.3 (0.6)	71.3 (0.7)	65.0 (1.0)
Cervix uteri	70.5 (1.6)	64.1 (1.8)	62.8 (2.1)	60.0 (2.4)
Corpus uteri and uterus, NOS	84.3 (1.0)	83.2 (1.3)	80.8 (1.7)	79.2 (2.0)
Ovary	55.0 (1.3)	49.3 (1.6)	49.9 (1.9)	49.6 (2.4)
Prostate	98.8 (0.4)	95.2 (0.9)	87.1 (1.7)	81.1 (3.0)
Testis	94.7 (1.1)	94.0 (1.3)	91.1 (1.8)	88.2 (2.3)
Urinary bladder	82.1 (1.0)	76.2 (1.4)	70.3 (1.9)	67.9 (2.4)
Kidney and renal pelvis	61.8 (1.3)	54.4 (1.6)	49.8 (2.0)	47.3 (2.6)
Brain and other nervous system	32.0 (1.4)	29.2 (1.5)	27.6 (1.6)	26.1 (1.9)
Thyroid	96.0 (0.8)	95.8 (1.2)	94.0 (1.6)	95.4 (2.1)
Hodgkin's disease	85.1 (1.7)	79.8 (2.0)	73.8 (2.4)	67.1 (2.8)
Non-Hodgkin lymphomas	57.8 (1.0)	46.3 (1.2)	38.3 (1.4)	34.3 (1.7)
Multiple myeloma	29.5 (1.6)	12.7 (1.5)	7.0 (1.3)	4.8 (1.5)
Leukaemias	42.5 (1.2)	32.4 (1.3)	29.7 (1.5)	26.2 (1.7)

The Lancet, 360 (October 12, 2002), 1131-1135


Digestive
Years 5 10 15 20
Rectum 63
Colon 62
Oral cavity, pharynx 57
55
54
52
52 Colon
49 Rectum
44
38
33
Stomach 24
19
19
15
Esophagus 14
Liver, bile duct 8
8
8
8
6
6
5
Pancreas 4
3
3
3
Respiratory
5 10 15 20
Larynx 69
57
46
38
Lung and bronchus 15
11
8
6
Reproductive
5 10 15 20
Prostate 99
Testis 95
95
94
91
88
87
Breast 86
Uterus 84
83
81
81
79
78
Cervix 71
71
65
64
63
60
Ovary 55
49
50
50
Other
5 10 15 20
Thyroid 96
96
94
95
Melanomas 89
87
Hodgkin's disease 85
84
83
Urinary, bladder 82
80
76
74
70
68
67
Kidney, renal pelvis 62
Non-Hodgkin's 58
54
50
47
46
Leukemia 43
38
34
Brain, nervous system 32
32
30
Multiple myeloma 30
29
28
26
13
7
5

Outline

Introduction

Principles of Data Visualisation
- Scale and range
- Theories of perception

Decoding: William Cleveland
- A hierarchy of visual decoding
  - Size (numeric)
  - Group (categorical)

Colour

Aesthetics: Tufte

Final words

Main Message

- Last two decades: new ideas on what is the most informative way to display quantitative information
  - Based on scientific knowledge on perception
  - Guidelines that help improving graph quality. “Mistakes” due to ignorance
  - No “right or wrong”, only “better or worse”

Main Message

- Last two decades: new ideas on what is the most informative way to display quantitative information
  - Based on scientific knowledge on perception
  - Guidelines that help improving graph quality. “Mistakes” due to ignorance
  - No “right or wrong”, only “better or worse”
- Graph creation iterative and laborious process, with trial and error.
  Creating a half-page figure takes same time as writing half-page text

Main Message

- Last two decades: new ideas on what is the most informative way to display quantitative information
  - Based on scientific knowledge on perception
  - Guidelines that help improving graph quality. “Mistakes” due to ignorance
  - No “right or wrong”, only “better or worse”
- Graph creation iterative and laborious process, with trial and error.
  Creating a half-page figure takes same time as writing half-page text
- Several high quality newspapers and magazines (New York Times, Financial Times, Washington Post, The Economist) are ahead of many scientists and institutions (including WHO)

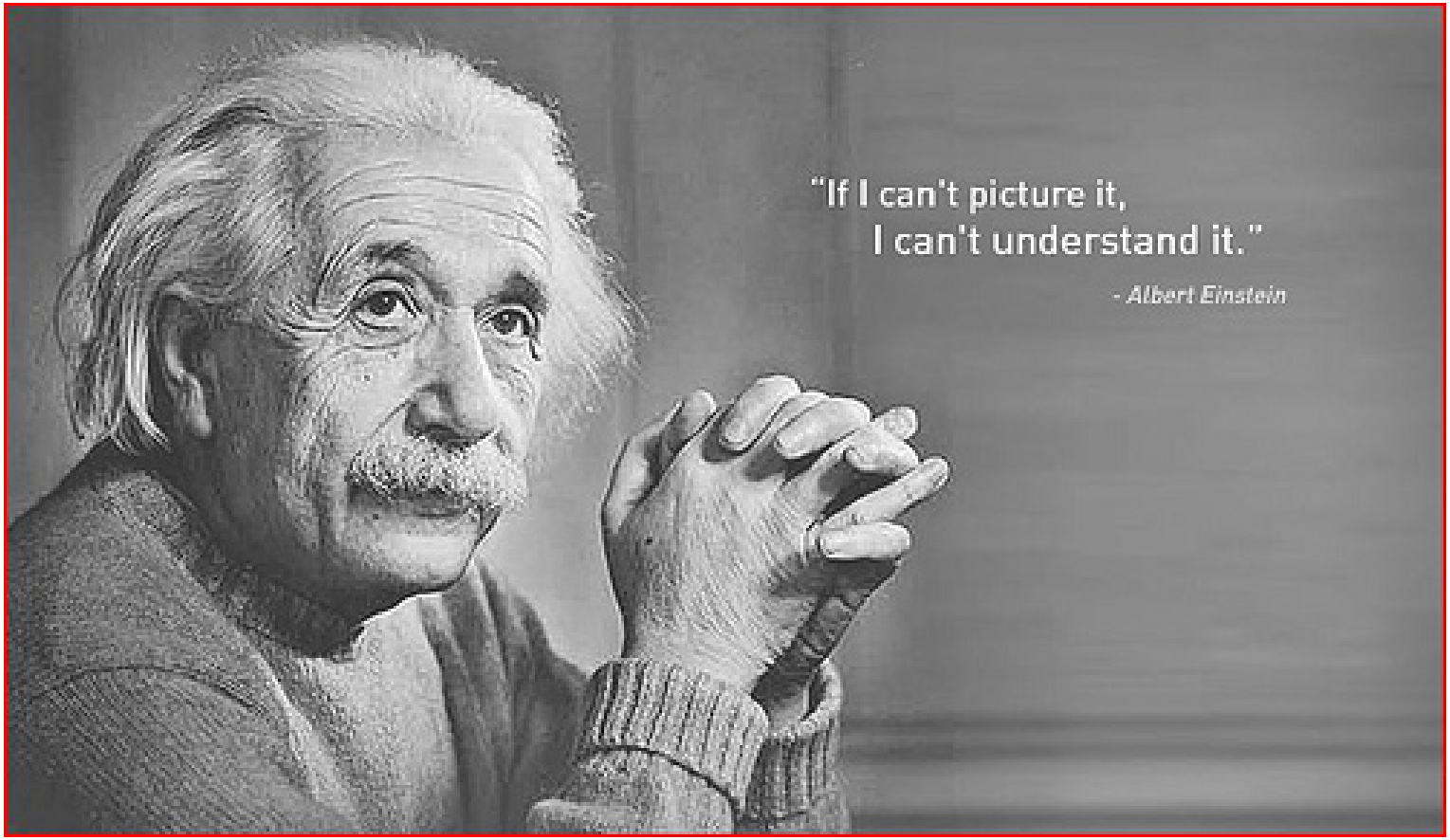
Graphics principles, the STRATOS initiative

From https://graphicsprinciples.github.io/threelaws.html

STRATOS

(STRengthening Analytical Thinking for Observational Studies)

visualisation panel


"If I can't picture it,
I can't understand it."
- Albert Einstein

Alberto Cairo.
The Truthful Art.
Pearson Education, 2016. http://www.thefunctionalart.com/

William S. Cleveland.
The Elements of Graphing Data.
Hobart Press, Summit, New Jersey, 1994.

Kieran Healy.
Data Visualization.
Princeton University Press, New Jersey, 2019. https://socviz.co/

Kate Strachnyi.
ColorWise.
O'Reilly, 2023.

Edward D. Tufte.
The Visual Display of Quantitative Information.
Graphics Press, Cheshire, Connecticut, 2001.

Klaus Wilke.
Fundamentals of Data Visualisation.
O'Reilly, 2019. https://serialmentor.com/dataviz/.