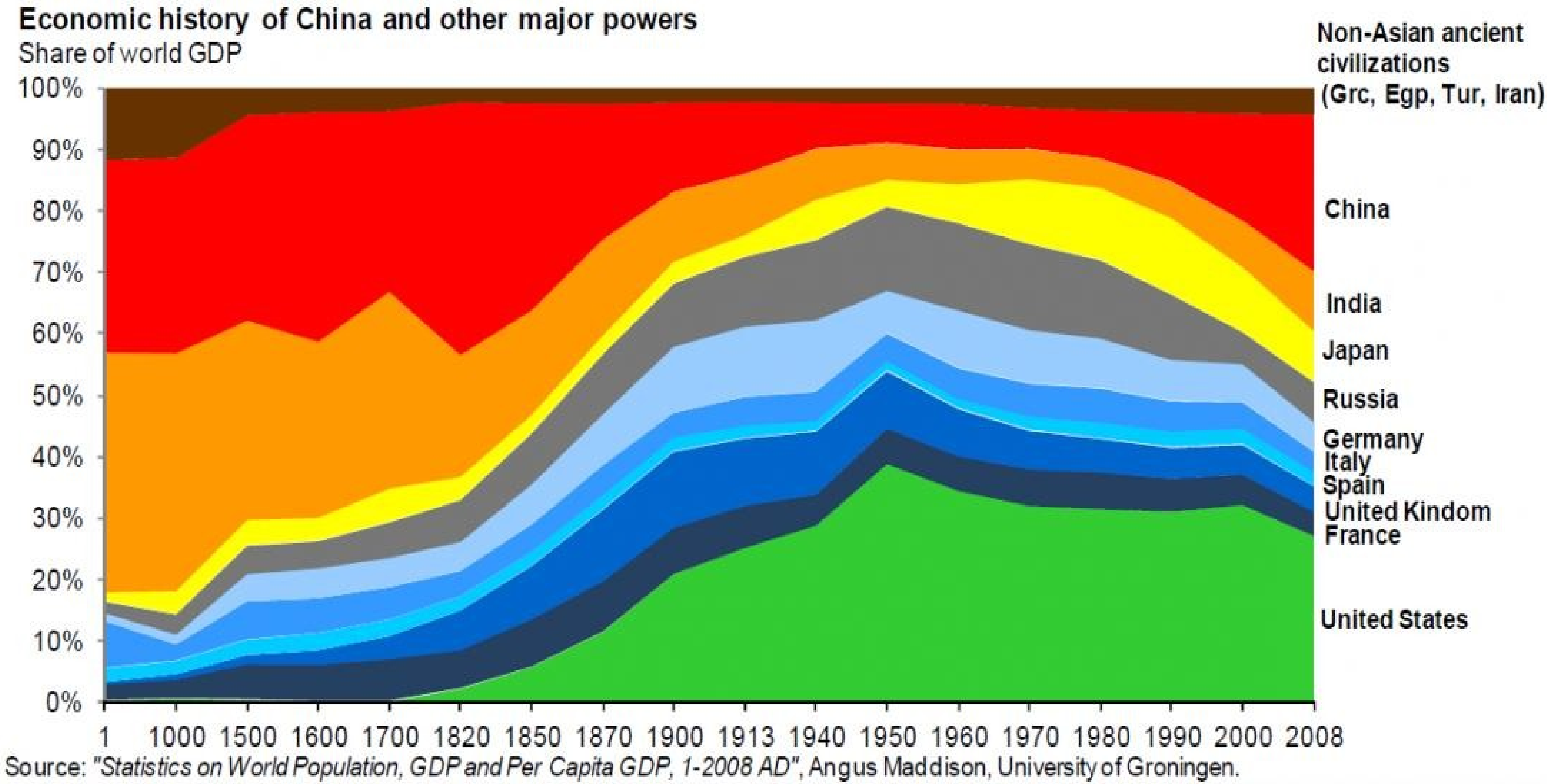
Data Visualisation
Lecture Week 8 - Encodings

Dr. Cathy Ennis

Learning Outcomes Week 8

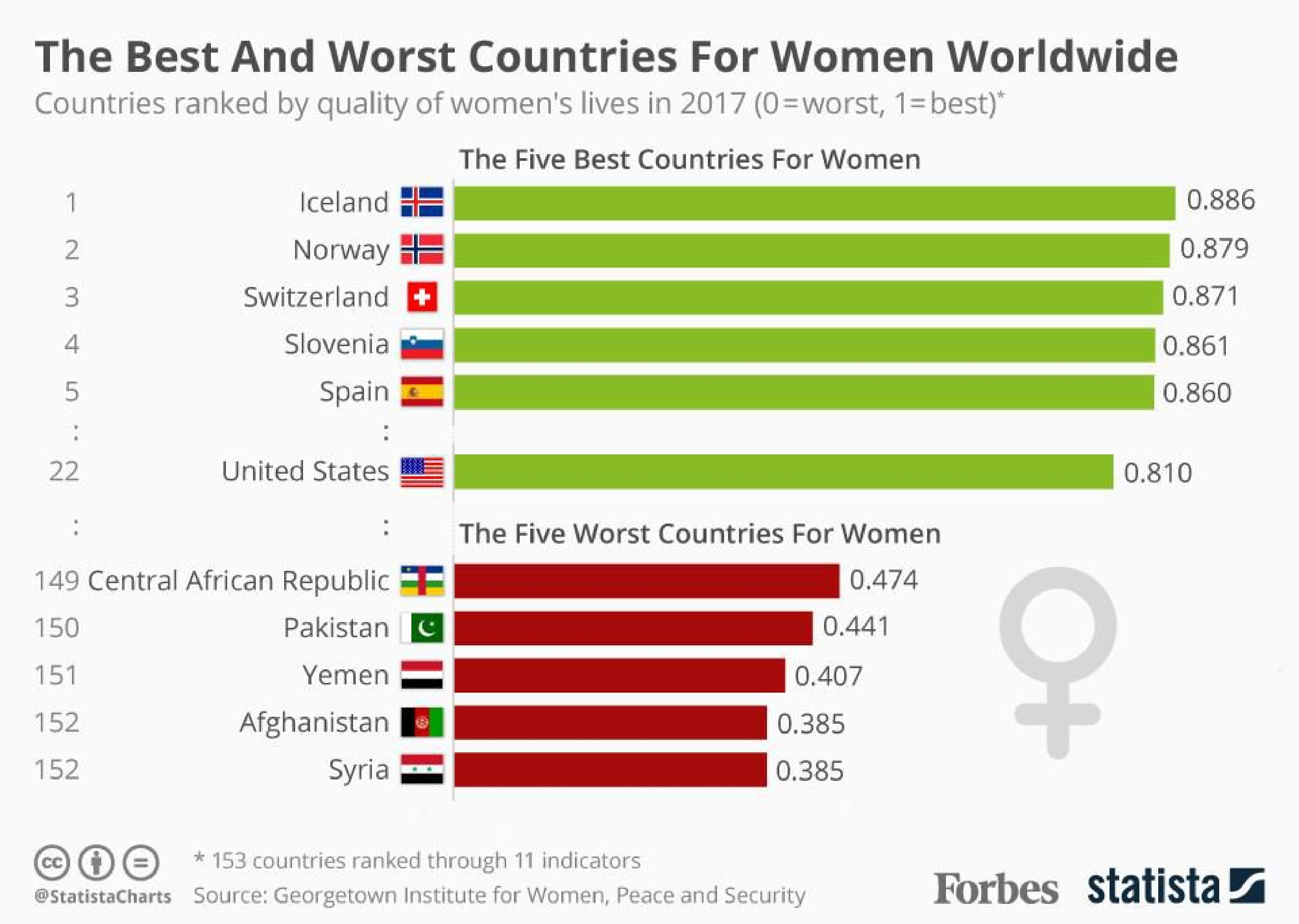
- Demonstrate understanding of how humans perceive the world around them on a general level and absorb complex data/information on a specific level.
- Analyse and evaluate how metaphors are used to convey unfamiliar information.
- Analyse and evaluate how mental models aid in the interpretation of complex visual displays.
- Select, formulate and integrate metaphors to suit data-driven tasks

Visualisation of the Week



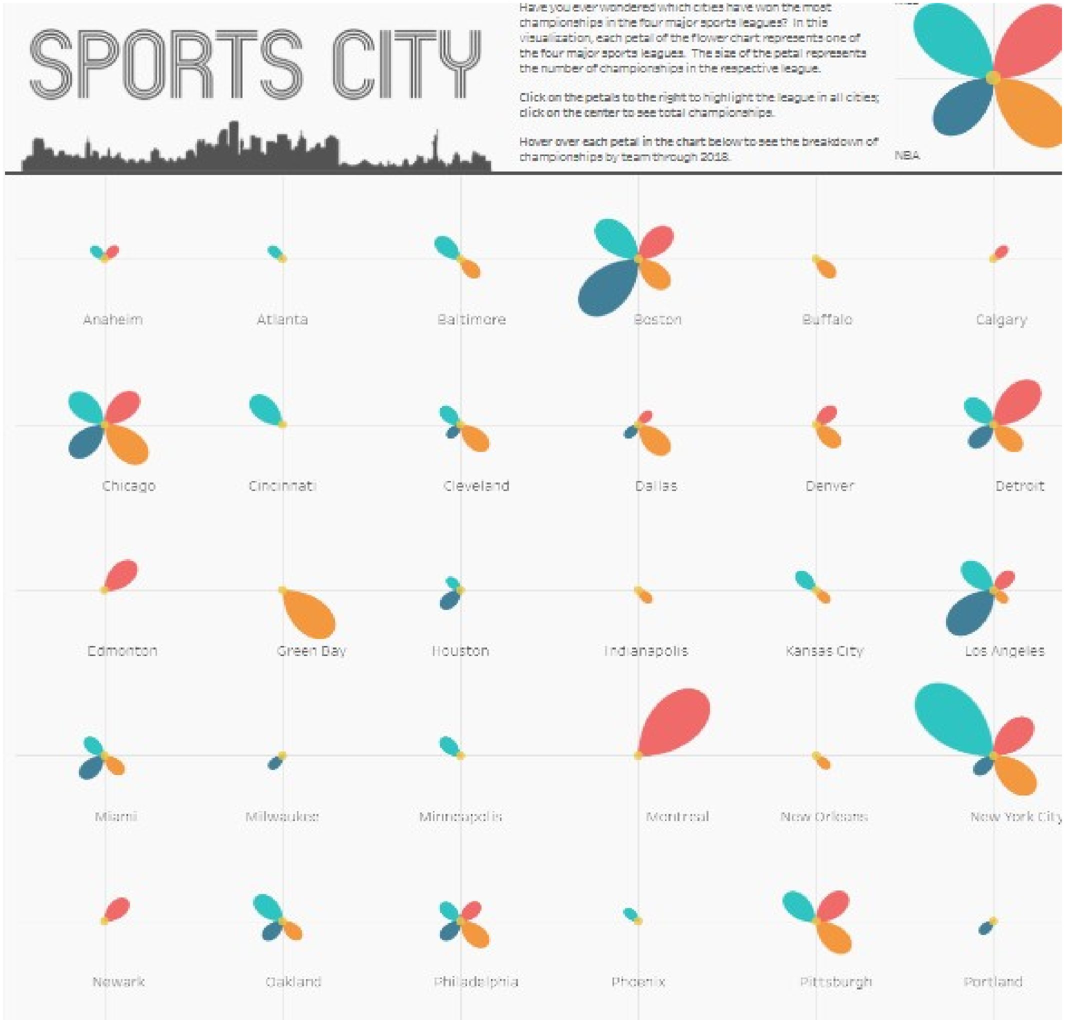
Visualisation of the Week

Data available: https://giwps.georgetown.edu/the-index/

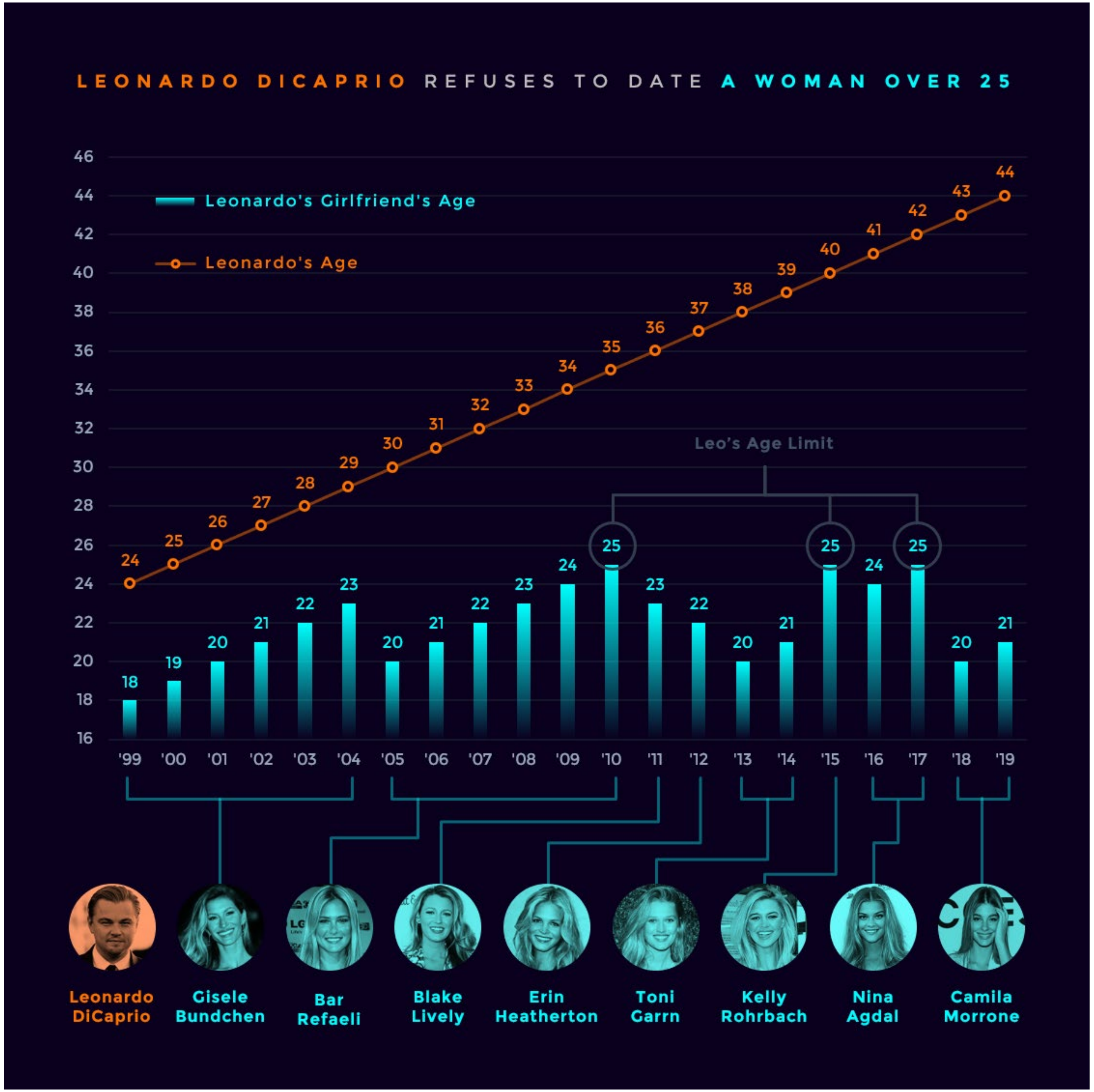


https://www.forbes.com/sites/niallmccarthy/2017/11/08/the-best-and-worst-countries-for-women-worldwide-infographic/

Visualization of the Week

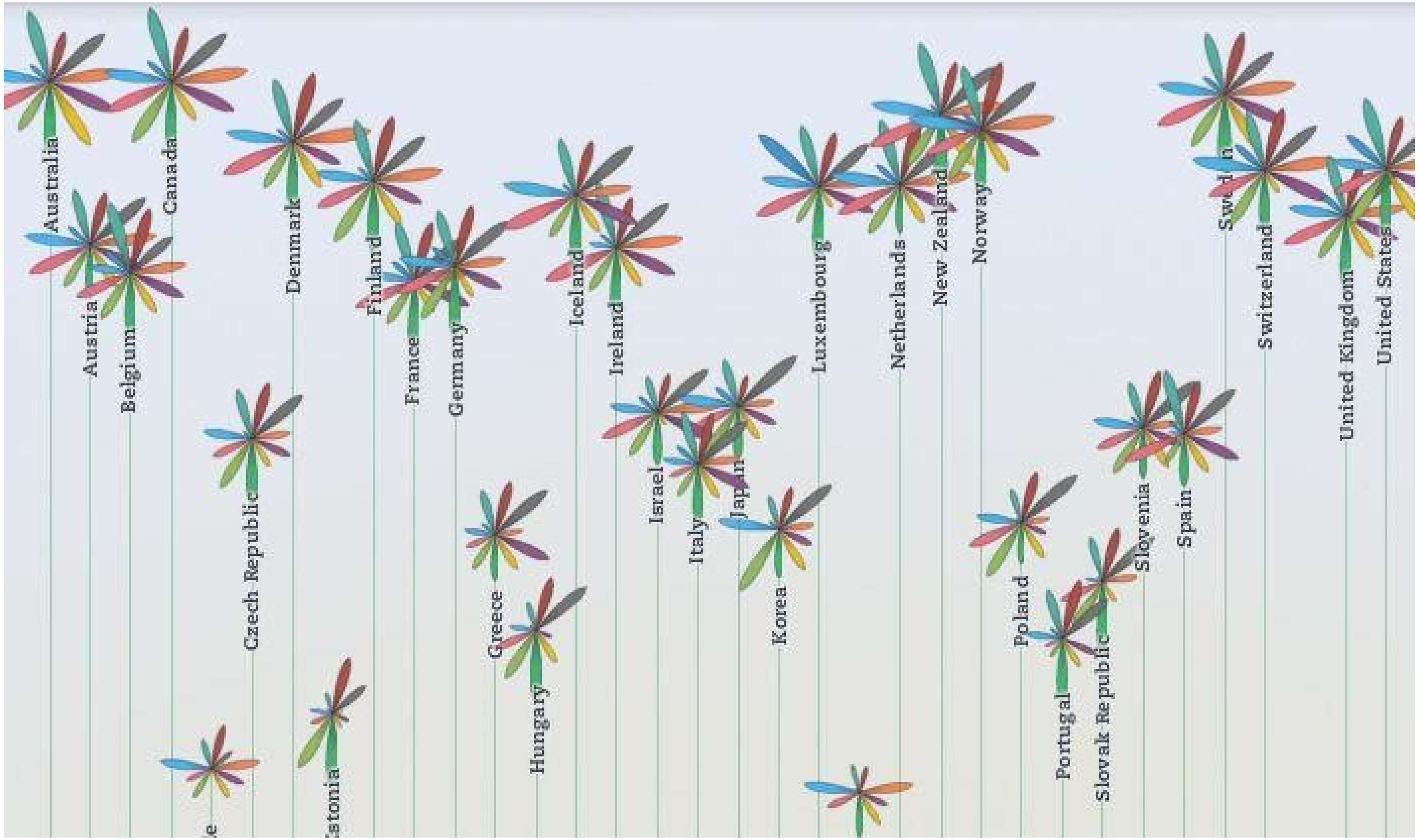


https://public.tableau.com/profile/matt.chambers#!/vizhome/SportsCity/SportsCity



https://www.reddit.com/r/dataisbeautiful/comments/azjti7/leonardo_dicaprio_refuses_to_date_a_woman_over_25/

(Un)Visualisation of the Week



Overview

- Data visualisations fundamentally map data dimensions to visual encodings
- In today's lecture we will look at the different encodings available and how we can appropriately use them

ENCODINGS

Data Types Classification

- Categorical (qualitative): no inherent order
  - city names, types of diseases, ...
- Ordinal (qualitative): ordered, but not at measurable intervals
  - first, second, third, …
  - cold, warm, hot
  - Mon, Tue, Wed, Thu …
- Quantitative: Integers or Reals
  - 1, 2, 7
  - 3.4, 78.6, - 12.4

Data Types

Categorical

- Text (qualitative) Nominal (qualitative)

Quantitative

- Interval
- Ratio

Ordinal

- Ordinal (qualitative)
- Temporal

Data Representation

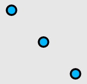
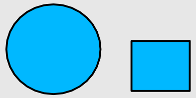
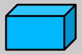
- Data representation is the act of giving visual form to your data
- Viewers decode the various shapes, sizes, positions and colours to form an understanding of the values presented
- Visualisers do the reverse through visual encodings, assigning visual properties to values

Data Representation

- Visual encoding of data involves:
  - Marks
  - Attributes

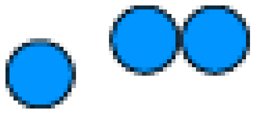
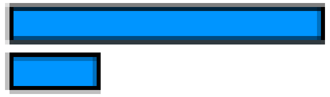
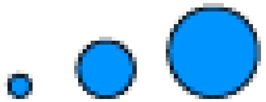
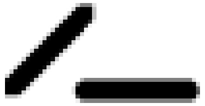
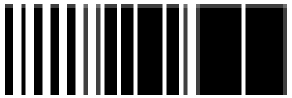
- The objective of visual encoding is to find the right blend of marks and attributes that most effectively will portray the angle of analysis you wish to show your viewers

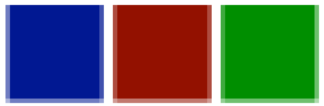
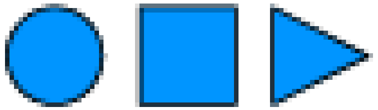
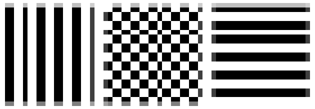
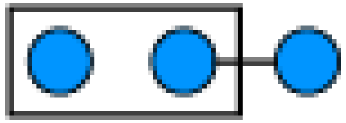
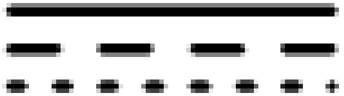
Marks

Point		**No variation**
Line		**One linear spatial dimension**
Area		**Two spatial dimensions**
Form		**Three spatial dimensions**

Attributes

Position
Size
Colour hue
Colour Saturation
Colour Lightness
Symbol/Shape
Pattern
Angle/Slope
…and others…

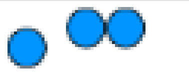
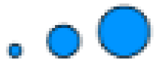
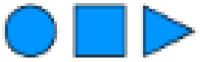
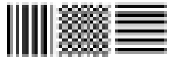
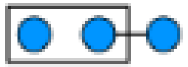
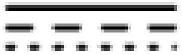
Example	Encoding	Ordered	Useful values
	position, placement	yes	infinite
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite
	length	yes	many
	size, area	yes	many
	angle	yes	medium
	pattern density	yes	few
	weight, boldness	yes	few
	saturation, brightness	yes	few

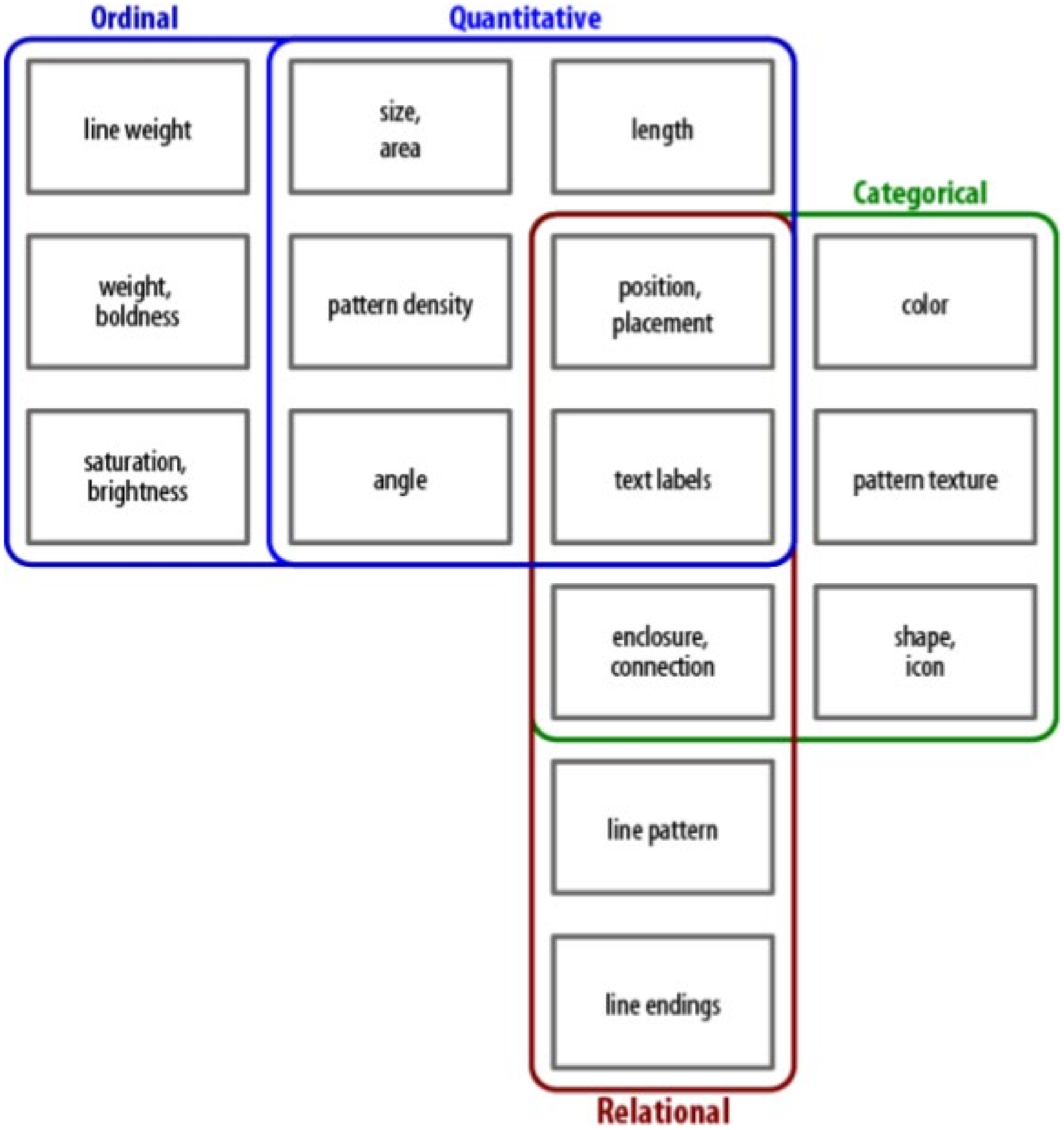
Example	Encoding	Ordered	Useful values
	color	no	few (<20)
	shape, icon	no	medium
	pattern texture	no	medium
	enclosure, connection	no	infinite
	line pattern	no	few
	line endings	no	few
	line weight	yes	few

Interpretations of Visual Encodings

Some properties have intrinsic meaning but some don’t:

- Density (Greyscale)
  - Darker -> More
- Size / Length / Area
  - Larger -> More
- Position
  - Leftmost -> first, Topmost -> first
- Hue
  - ??? no intrinsic meaning

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		


Ordinal
Quantitative
line weight
size,
area
length
Categorical
weight,
boldness
pattern density
position,
placement
color
saturation,
brightness
angle
text labels
pattern texture
enclosure,
connection
shape,
icon
line pattern
line endings
Relational

Appropriate Visual Encodings

Things to think about when choosing Appropriate Visual Encodings:

- Natural ordering
- Distinct values
- Defaults versus innovative formats
- Readers’ context
- Compatibility with Reality
- Patterns and Consistency

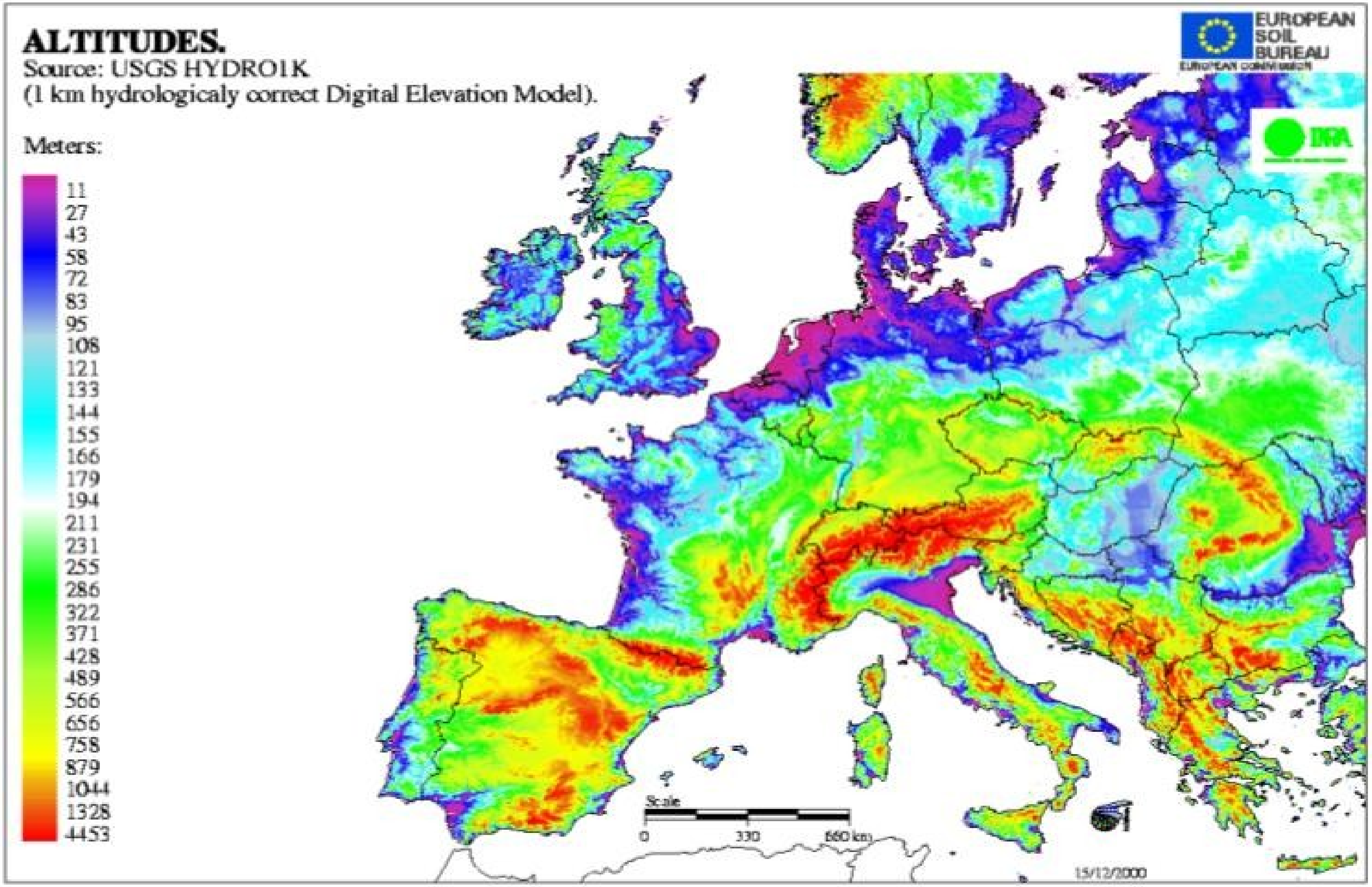
Natural Ordering

- When there is an order or ranking assigned to values of a visual property
  - Determined subconsciously by the mechanics of our visual system
- Process is deeply embedded in our understanding and evaluates relative order independent of language, culture or convention
  - Automatic and unintentional
- Depending on the specifics of the visual property, its natural ordering may be well suited to representing:
  - Quantitative differences
  - Ordinal differences

Examples of Natural Ordering

- Position has a natural ordering
  - shape does not
- Length has a natural ordering
  - texture does not (but pattern density does)
- Line thickness or weight has a natural ordering
  - line style (solid, dotted, dashed) does not

Colour Is Not Ordered



Colour Is Not Ordered

- Colour (hue) is not naturally ordered in our brains
  - brightness (lightness or luminance, sometimes called tint) is
  - intensity (saturation) is

- Strong **social conventions** about colour

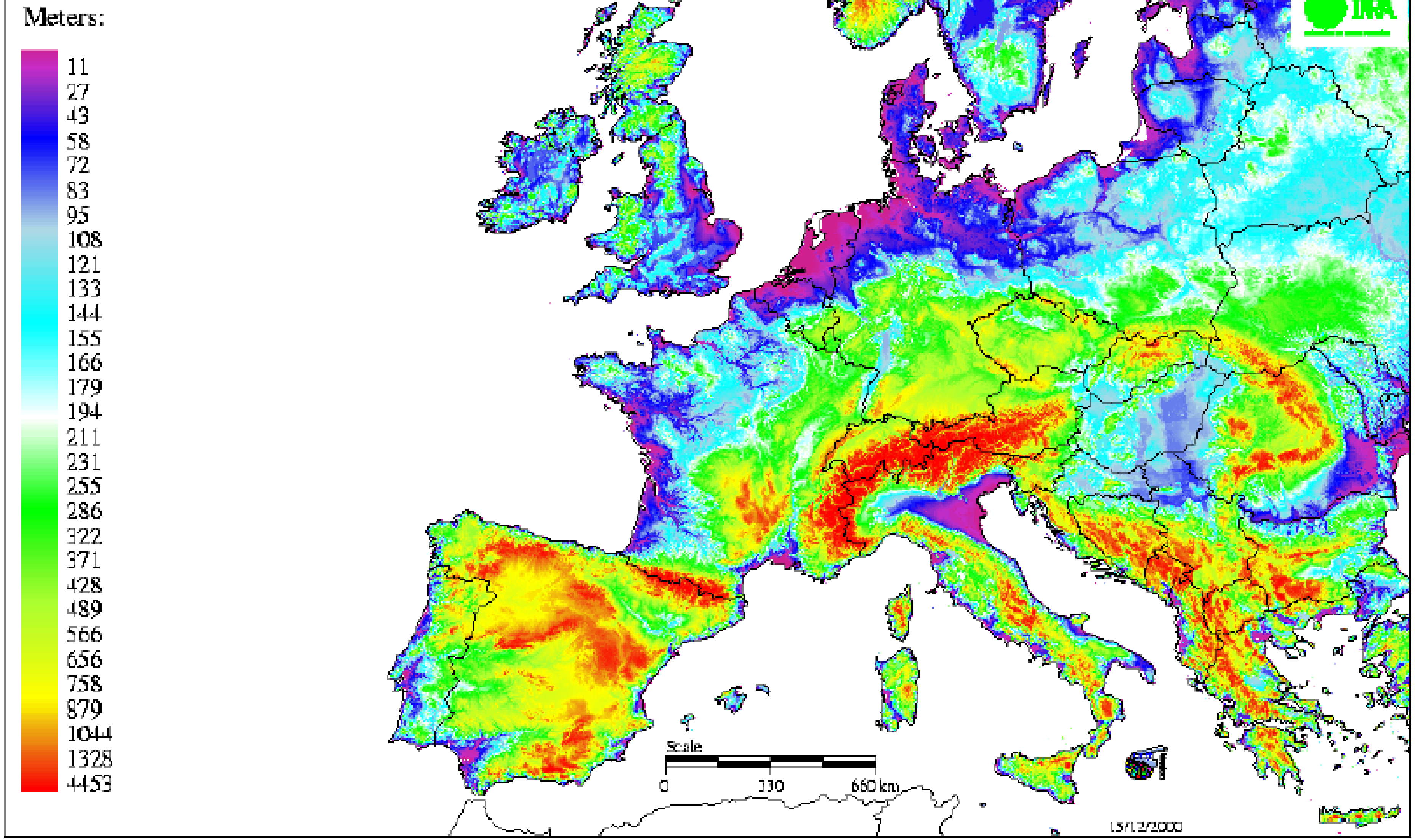
Misuse of Colour

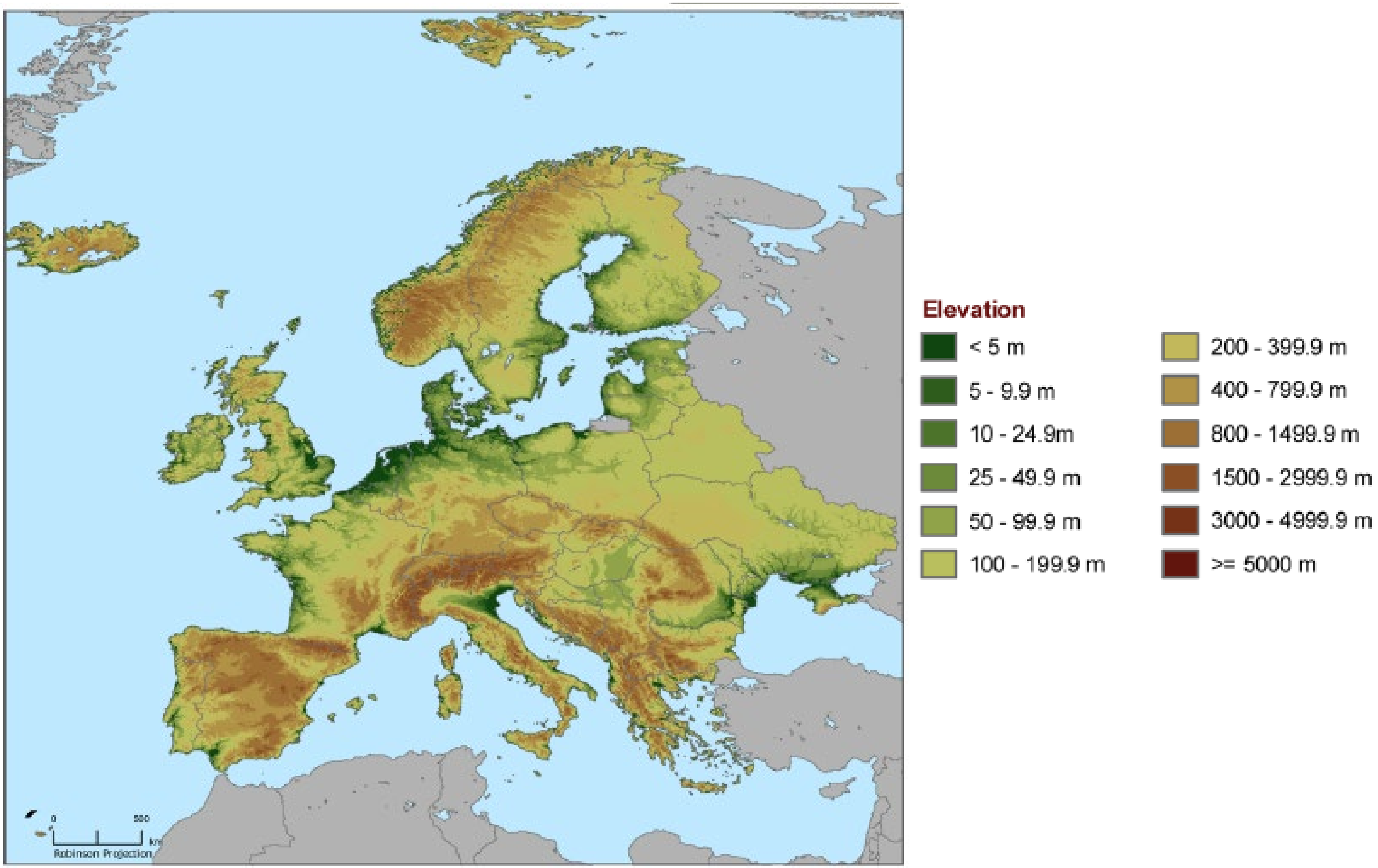
- Misuse of colour to imply order is rampant
- If tempted to use **ordered colour** (elevation, heat maps, etc.), consider varying brightness/Saturation
- For example, elevation can be represented by increasing the darkness of browns, rather than cycling through the rainbow

ALTITUDES.

Source: USGS HYDRO1K

(1 km hydrologicaly correct Digital Elevation Model).




Elevation
< 5 m
5 - 9.9 m
10 - 24.9m
25 - 49.9 m
50 - 99.9 m
100 - 199.9 m
200 - 399.9 m
400 - 799.9 m
800 - 1499.9 m
1500 - 2999.9 m
3000 - 4999.9 m
>= 5000 m
0
500
Robinson Projection

Distinct Values

- When choosing a visual property, select one that has a number of useful differentiable values and an ordering similar to that of your data

- There are a lot of colours in the world, but we cannot tell them apart if too similar

- We can more easily differentiate a large number of shapes, a huge number of positions, and an infinite number of numbers

Distinct Values

Example	Encoding	Ordered	Useful values
	position, placement	yes	infinite
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite
	length	yes	many
	size, area	yes	many
	angle	yes	medium
	pattern density	yes	few
	weight, boldness	yes	few
	saturation, brightness	yes	few
	color	no	few (<20)
	shape, icon	no	medium
	pattern texture	no	medium
	enclosure, connection	no	infinite

Defaults vs Innovative Formats

- There are a lot of good default encodings and encoding conventions in the world, with good reason:
  - Designing new encoding formats can cost you a lot of time and effort
  - Knowing the expected defaults for your industry, data type, or reader can save you a lot of work when it comes to figuring out how to best encode your data

Defaults vs Innovative Formats

- Sometimes default formats can put readers to sleep, but other times it helps them get to the heart of the matter

- Cost/Benefit:
  - If you have a superior solution then use it, but if your job can be done well enough with a default format, save everyone the effort and use a standard solution

Readers’ Context - Audience

- You are creating a visualization for someone other than yourself
    - The reader may have a different mindset or view the world differently to you
    - Your audience will likely be composed of more than one reader
- Choose the most important group, think of them as your core group, and design with them in mind
    - Where possible to appeal to more people without sacrificing precision or efficiency, do so
- “Reader” = a representative reader from within your core audience

Readers’ Context – Convention

- Scientific universality about how we perceive and process colour but also significant cultural associations
  - some colours may be lucky, some unlucky; some may carry positive or negative connotations; some may be associated with life events
  - Colours may also take on special significance when paired with certain shapes - a red octagon means stop in many places
  - Others take on more significance when paired/grouped with other colours

This stop sign from Montreal is labelled in French, but no English speaker is likely to be confused about its meaning.

Readers’ Context - Colour Blindness

- Many variations in the way different people perceive colour commonly called “colour blindness” but is more properly referred to as “colour vision deficiency” or dyschromatopsia
- About 7-8% of men experience some kind of colour perception disorder. Women are much more rarely affected: about 0.4%.
- Red-green deficiency is the most common by far, but yellow-blue deficiency also occurs, also close colours like blue and purple
- https://www.iamcal.com/misc/colors/

Readers’ Context - Culture

- Is the reader from a culture that reads left-to- right, right-to-left, or top-to-bottom?
  - A person’s habitual reading patterns will determine their default eye movements over a page, and the order in which they will encounter the various visual elements in your design
- It will also affect what the reader perceives as “earlier” and “later” in a timeline
- May also pertain to geographic maps: many of us are used to seeing the globe split somewhere along the Pacific, with north oriented upward
- Colour associations in different cultures: https://informationisbeautiful.net/visualizations/colours-in-cultures/

https://cartorical.com/shop/pacific-centered-world-map/

Compatibility With Reality

- Make your encodings of things and relationships as well aligned as possible with the reality (or your reader's reality) of those things and relationships
  - Some things are larger than others, have specific colours, well-known locations, and other identifying characteristics

- This alignment is called **compatibility**

Compatibility With Reality

- Avoid countering conventions where possible to avoid creating **cognitive dissonance**
  - Clash of habitual interpretation with the underlying message you are sending

- Consider cultural conventions about spatial representations, such as what left and right mean politically, or the significance of above and below

Patterns & Consistency

- As designers, we must be extremely deliberate about the patterns and pattern violations we create
- Three simple rules:
  - Be consistent in membership, ordering, and other encodings
  - Things that are the same should look the same
  - Things that are different should look different
- Maintaining consistency and intention when encoding will greatly enhance the accessibility and efficiency of your visualization

Appropriate Visual Encodings

- Choosing Appropriate Visual Encodings
  - Natural ordering
  - Distinct values
  - Defaults versus innovative formats
  - Readers’ context
  - Compatibility with Reality
  - Patterns and Consistency

Exercise 1 - R

- An analysis of student grades with R
- In this lab, you will use ggplot and sqldf to analyse student marks in a very small file.

Name	DOB	Subject	Year	Grade	Mark_Written	Mark_Oral
Mary Healy	10-06-1988	Maths	2015	A	82	79
Mary Healy	10-06-1988	English	2015	A	84	78
Mary Healy	10-06-1988	Irish	2015	A	87	76
Mary Healy	10-06-1988	Japanese	2015	A	99	90
Mary Healy	10-06-1988	Chinese	2015	A	98	93
Joe O'Neil	4-03-1979	Maths	2015	B	76	70

ggplot structure

myplot <- ggplot(data= yourdataset, aes(x=yourx, y =youry))
#this begins your plot by adding the data

<u>Examples of extra layers</u>

myplot +geom_point() #this adds a geometry to your plot (scatter plot in the example)

myplot +geom_point(aes(colour=dimension)) #geom layer can be customized

myplot +geom_bar()+scale_fill_brewer(palette='Reds') #customizing the colour palette

myplot +geom_bar(colour=dimension) +scale_fill_manual(values=c('blue','red'))

#customize colours manually

ggplot structure

- **myplot <- ggplot(data= yourdataset, aes(x=yourx, y =youry))**
- #this begins your plot by adding the data

- <u>Examples of extra layers</u>

- myplot+coord_map(projection="ortho", orientation= c(41,-74,0)) #map projection

- myplot +theme_classic() #applies a predefined theme to the plot

- myplot +labs(title="graph title", x="xaxistitle", y="yaxistitle") #labels

- myplot +facet_wrap(dimension) #creates small multiples based on dimension

An analysis of student grades with R Examples

- Inspecting data, Dealing with missing values.
- Question 1. What are the average results in written exams across all subjects and all years per student?
- Question 2. What are the average results in oral exams across all subjects and all years per student?
- Question 3. What are the average results in the written exams per student and year?
- Creating custom functions.

An analysis of student grades with R Exercises

1. What are the total marks (oral plus written divided by two) for each student for each subject? (2 marks)

2. What is the relationship between age and mark? (2 marks)

3. Did any students do better on their written compared with their oral (or vice versa)? (2 marks)

4. What subject obtained the best results on average? (2 marks)

5. What are the average results in oral exams across all subjects and all years per student? (2 marks)

Thanks To

- Marisa Llorens-Salvador, John McAuley, Colman McMahon and Brian Mac Namee for an earlier version of these lecture notes