# Geospatial Data Analytics for Business 1

Giorgio Chiovelli, Sebastian Hohmann

19/03/2020

# Table of contents

# Introduction

## What are these sessions about?

**Four aims**

- Give you a (selective) overview of how *Geographic Information Systems* (GIS) methods are used Business.
- Introduce you to QGIS software and (optionally) GIS scripting in python.
- Show you how you can "do GIS" for business through two use cases.

**We want to make this useful for you!**

- If you bring your laptop with QGIS installed, you can follow what we are doing in class.
- The course is supposed to be interactive – please ask questions, give us comments!

**Material for exercises**. Google drive

**Q&A**. Google doc

## What we will do in each session

### Today (19/03/2020)

- Overview of GIS applications in Business
- Introduction to QGIS, setup, simple examples
- Use Case 1: Spatial Analysis for Real Estate

### Tomorrow (20/03/2020)

- Use Case 2: Optimal store location
- Introduction to python scripting
- Python implementation of Use Case 2

**Let's keep our terms straight**

- **GIS** – Geographic Information Systems – systems to read, visualize, analyze, transform, and manage geospatial data.
- **geospatial data** regular data but with two additions
  1) co-ordinates ($x, y$ (more rarely $z$)), most commonly latitude-longitude
  2) a co-ordinate reference system that allows locating pieces of data vis-a-vis each other
- **Why do we care?**

# Introduction

## Why use GIS with analytics for Business?

**1) Locating business data**

- Everything – transactions, consumers, comptetitors, assets,... – is somewhere
- Attaching geographic information to a piece of data we already have $\rightarrow$ increase granularity, ability to predict and make decisions based on the same data

**2) Analyzing movement**

- Repeatedly observing the location of objects $\rightarrow$ can analyze change and make predictions

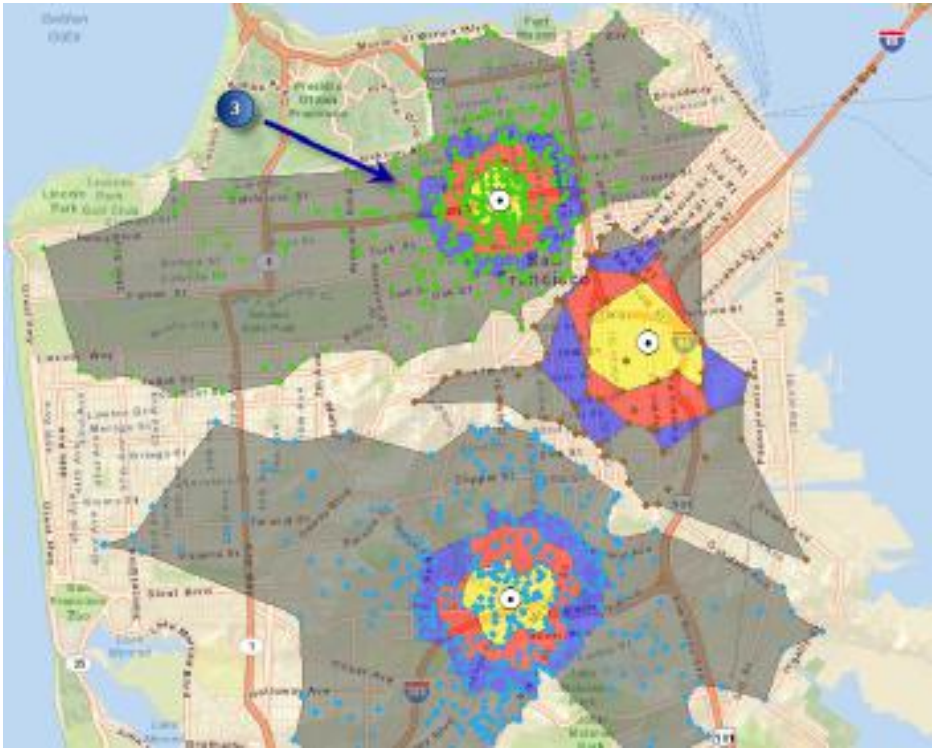
**3) Visualizing data**

- Putting data on a map is powerful. Done right, this can
  - deliver insights.
  - convince clients.

# Examples of Business GIS use cases

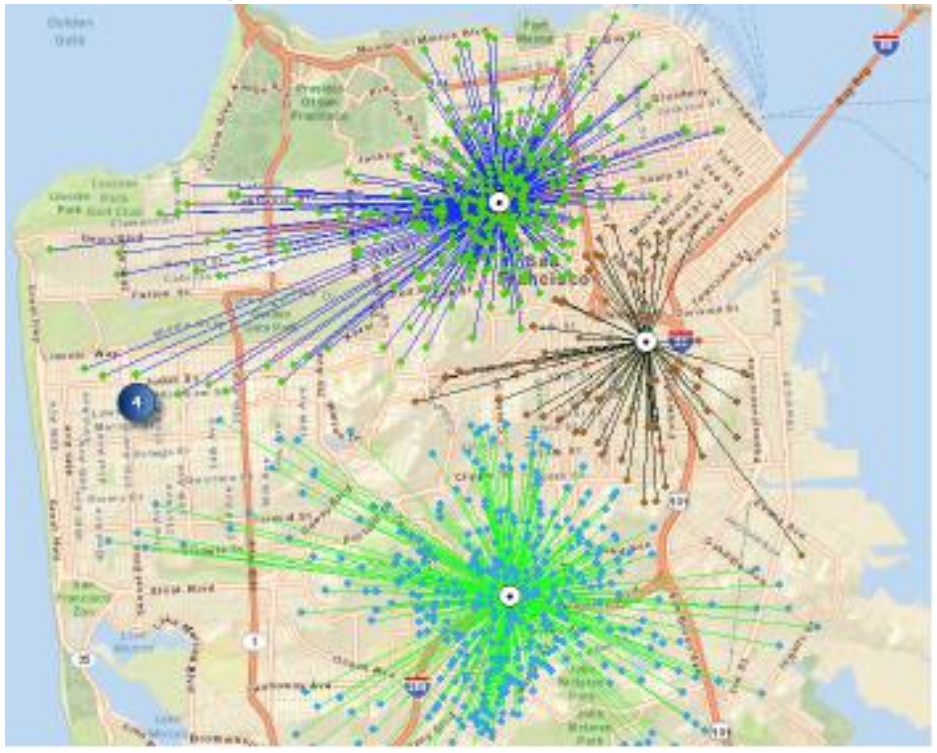
## Retail 1: Location Allocation

Market share areas



source: `http://desktop.arcgis.com/en/arcmap/latest/extensions/business-analyst/location-allocation.htm`

Spider diagram



Each demand point associated with allocated facility via straight lines

# Examples of Business GIS use cases

## Retail 1: Location Allocation

**Business:** Chain store

**Goal:** Choose the best location for a new branch among a set of "candidates"

- where do I want to set up?
- where are my customers? my competitors?
- how will my customers travel to my store?

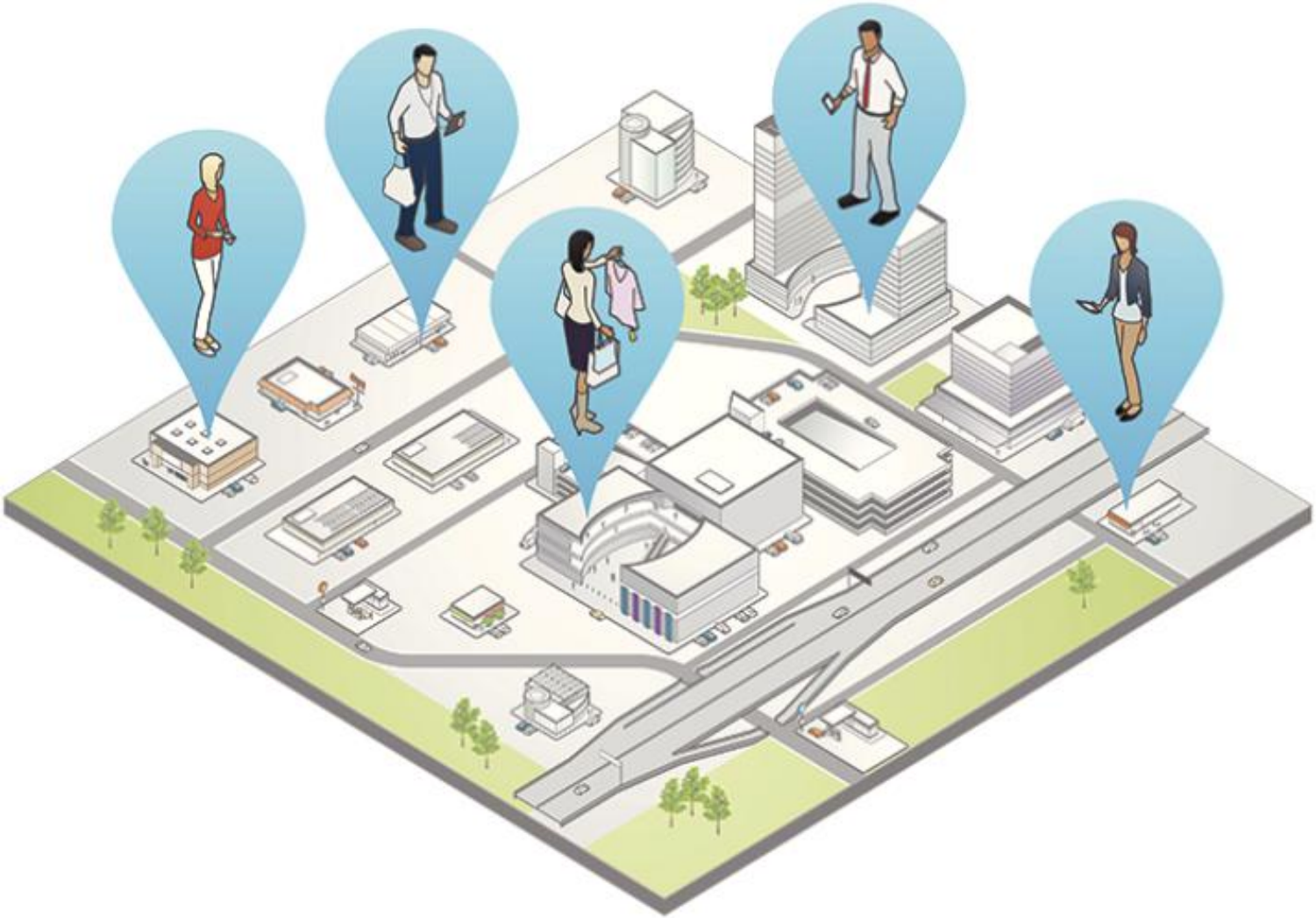
**How GIS can help:** Combine (at least) three types of data to find the stores with most consumers.

- set of possible store locations
- demands points (representing number of people, their incomes...)
- network dataset to model travel times between locations and demand points (or simple distance)

**Examples of firms:** *Establish (supply chain consulting firm), Petco (pet specialty retailer)*

# Examples of Business GIS use cases

## Marketing 1: More precise ad-targeting of consumers



source: Getty Images

# Examples of Business GIS use cases

## Marketing 1: More precise ad-targeting of consumers

**Business:** Telecom company

**Goal:** Send ads to customer phones, but not

- on Monday morning.
- when it is raining.
- when the customers are sitting in public transport.

**How GIS can help:** Mobile-phone location data enables firms to target consumers when and where they are more receptive.

**Examples of firms:** *Foursquare*, *xAd* (location marketing companies)

# Examples of Business GIS use cases

## Marketing 1: More precise ad-targeting of consumers



source: xAd

- Location based targeting of individual consumers creates enormous potential for advertisers
- Be aware of ethical and compliance issues!
  1) Data collection
  2) Data useage
  3) Data sale

`https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html`

# Examples of Business GIS use cases

## Marketing 2: Building audiences


source: xAd

# Examples of Business GIS use cases

## Marketing 2: Building audiences

**Business:** Various

**Goal:** Find out about audiences: work, habits, preferences

- Where do they live?
- Where do they work?
- What do they do in their free time?
- What shops do they visit?

**How GIS can help:** Mobile-phone location data enables firms to segment audiences.

To be useful, location data must

1) see people repeatedly and often
2) be able to pinpoint locations precisely. For wide fence, 80% of visitors seen inside fence don't ever go into business of interest. "Geo-fencing" can be of limited usefulness if want to build audiences based on real world visitation behavior.


**Examples of firms:** *xAd*

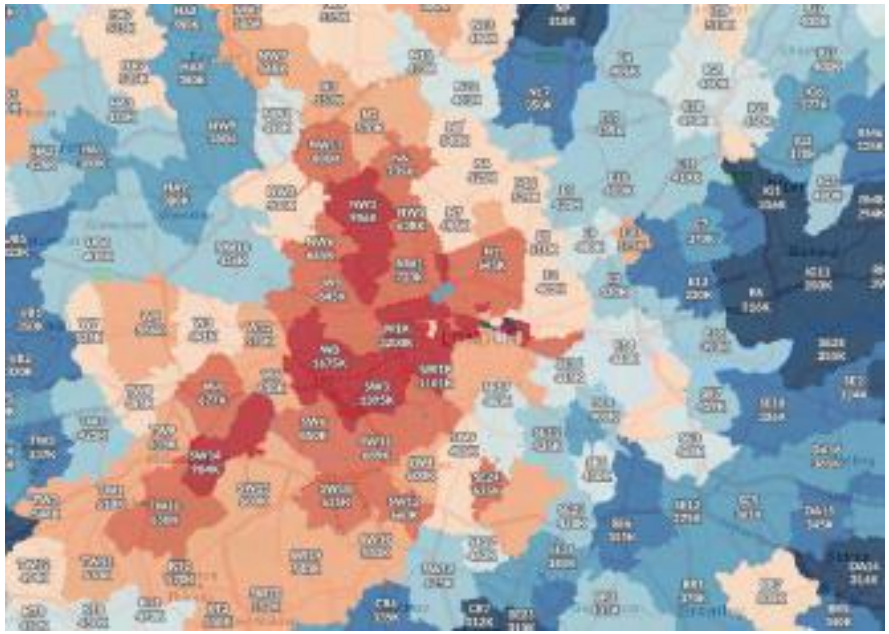
# Examples of Business GIS use cases

## Real Estate 1: Visualization of housing market and neighbourhood characteristics

antisocial behavior

source: `illustreets.com`

house prices



# Examples of Business GIS use cases

## Real Estate 1: Visualization of housing market and neighbourhood characteristics

**Business:** Estate agent

**Goal:** Present information on neighbourhoods to potential clients

- average home values ( *"Do I have a great deal for you!?"*)
- crime rates
- average ages and incomes
- presence of amenities (schools, parks, cafes, public transport)
- ...

**How GIS can help:** Putting the data on the map and visualizing them is powerful when communicating a lot of information quickly.

**Examples of firms:** *illustreets*

# Examples of Business GIS use cases

## Logistics 1: Avoiding Left Turns

In 2004, UPS instituted a "no left turn" policy for its delivery truck drivers


source:
`https://wonderfulengineering.com/dont-ups-drivers-turn-left/`

route with left turns

route without left turns

source: Sygic Navigation

# Examples of Business GIS use cases

## Logistics 1: Avoiding Left Turns

**Business:** Logistics Company

**Goal:** Avoid / minimize the amount of left turns (right in the UK) drivers make when delivering packages

- turning left = turning into oncoming traffic
- have to wait when turning left
- longer delivery times, higher petrol consumption, more accidents

**How GIS can help:** plan routes optimally, avoid left turns among other things.

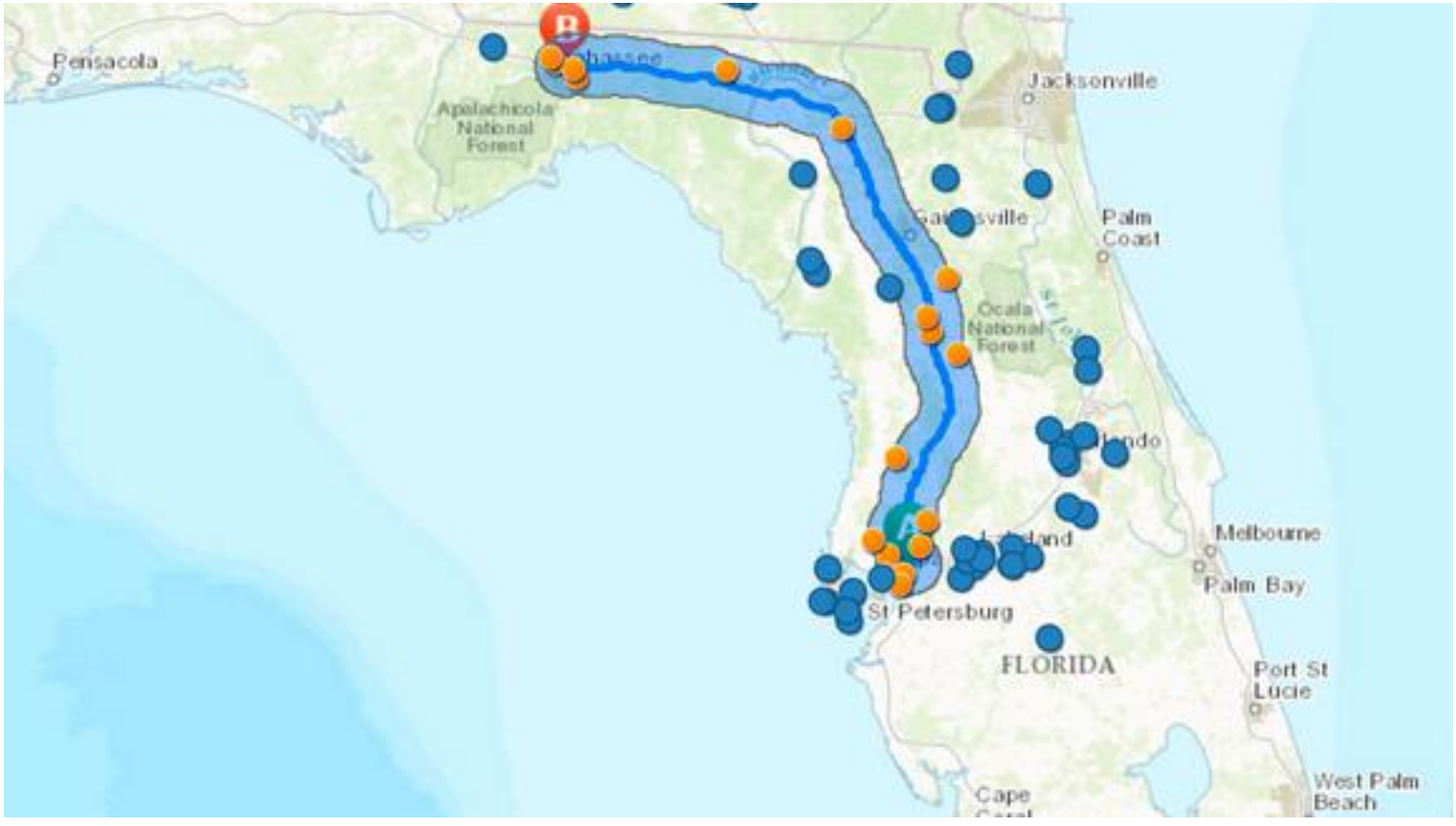
**Examples of firms:** *UPS*: instituted right-turn policy in 2004, routes consist of *approx* 90% RH-turns, firm claims has allowed it to

- use 10m gallons less fuel
- emit 20k tons less CO2
- cut fleet of trucks by 1100
- cut total distance travelled by its trucks by 28.5m miles despite longer routes on average

# Examples of Business GIS use cases

## Logistics 2: Fraud detection

Delivery route and gas purchases by sub-contractor




source: `https://www.esri.com/about/newsroom/publications/wherenext/secret-weapon-fraud/`

# Examples of Business GIS use cases

## Logistics 2: Fraud detection

**Business:** Logistics Company

**Goal:** Detect fraud by sub-contractors

- Deliveries carried out by sub-contractor
- Given gas card to buy fuel
- Unusually high gas consumption

**How GIS can help:** Plot routes and location of gas purchases on map.

- purchases far outside the route immediately apparent

# Examples of Business GIS use cases

## Insurance 1: Home Insurance

**Aerial photographs can be used to assess insurance risks**


source: Cape Analytics

# Examples of Business GIS use cases

## Insurance 1: Home Insurance

**Business:** Insurance Company

**Goal:** Price property damage risk more accurately

- 40% of all home losses are in some way related to the roof
- traditional data sources used by insurers to price risk
  - insurance agents:
  - public records (tax records, building permits): often inaccurate, outdated
  - inspections: accurate but expensive to collect, often available only months after policy is bound
- Insurance companies always seeking better information on the build environment: trillions of Dollars at stake!

**How GIS can help:** Aerial photos are objective, timely and comprehensive datasource for roof characteristics and conditions

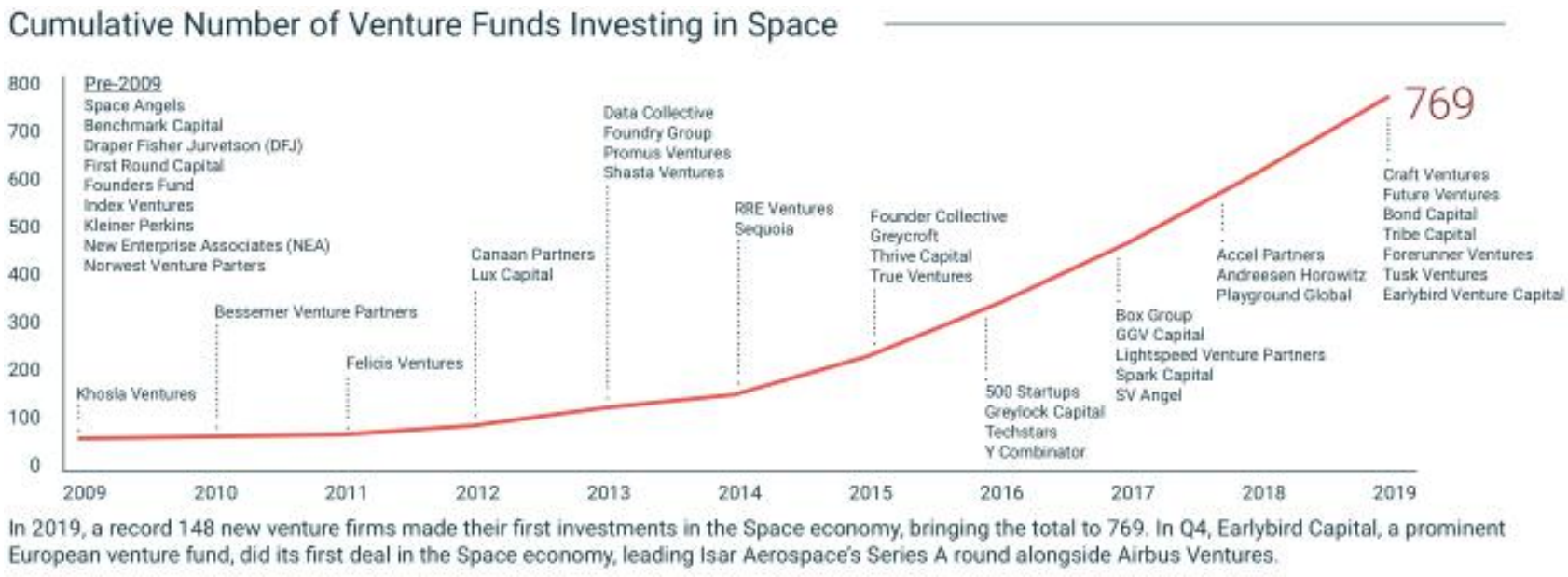
**Examples of firms:** *Cape Analytics (US), Landmark (UK)*: obtain aerial photographs and other geospatial imagery from their partners, provide API to insurance companies like *XL Catlin* who send lat/lon or address queries and receive structured data

- roof condition ("good", "bad",...), roof geometry ("flat", "pitched",...), roof covering ("tile", "tar",...), solar panels ("yes", "no"), ...

# Examples of Business GIS use cases

## Satellites

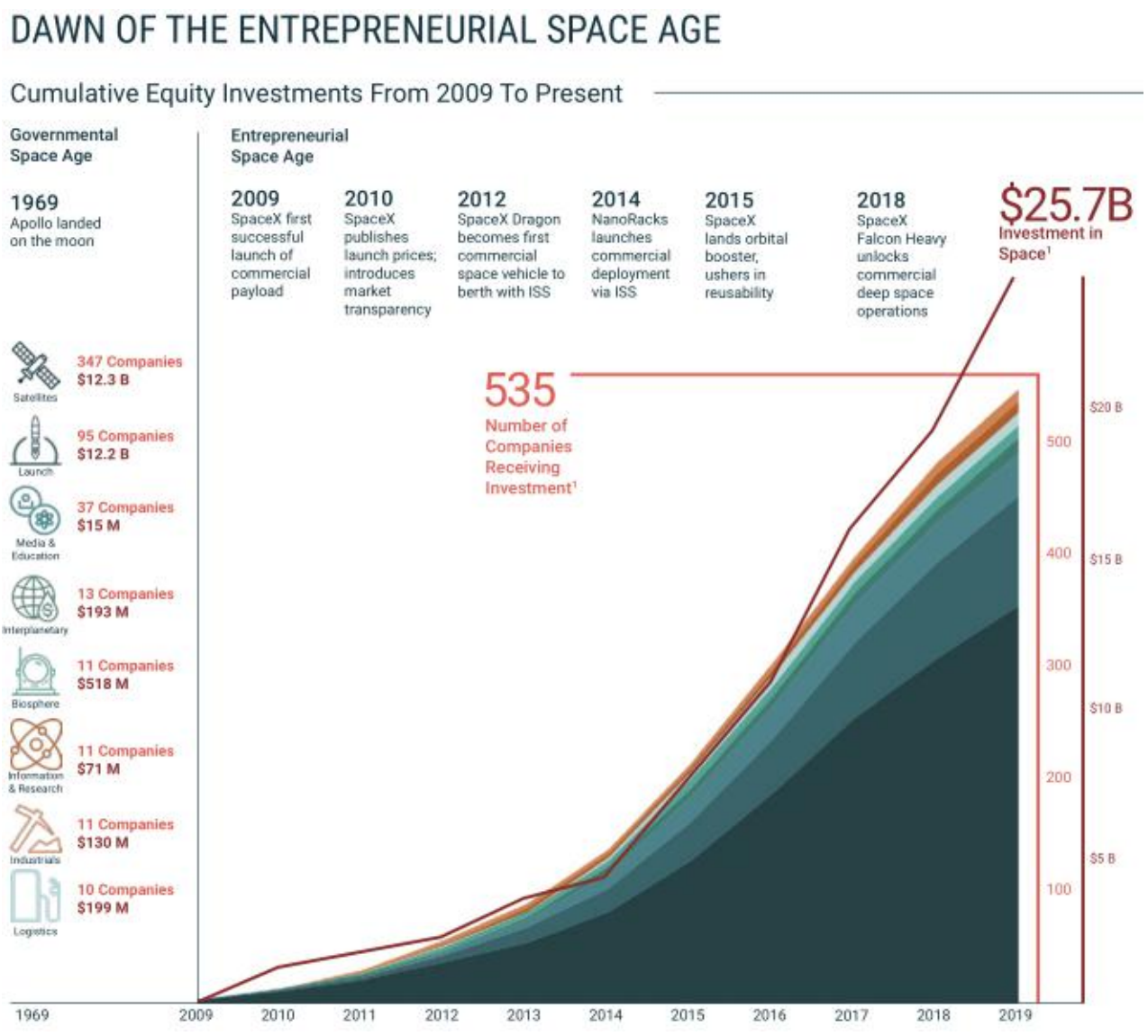
One set of GIS applications has exploded thanks to private investment in satellite technology




source: Space Angels

# Examples of Business GIS use cases

## Satellites

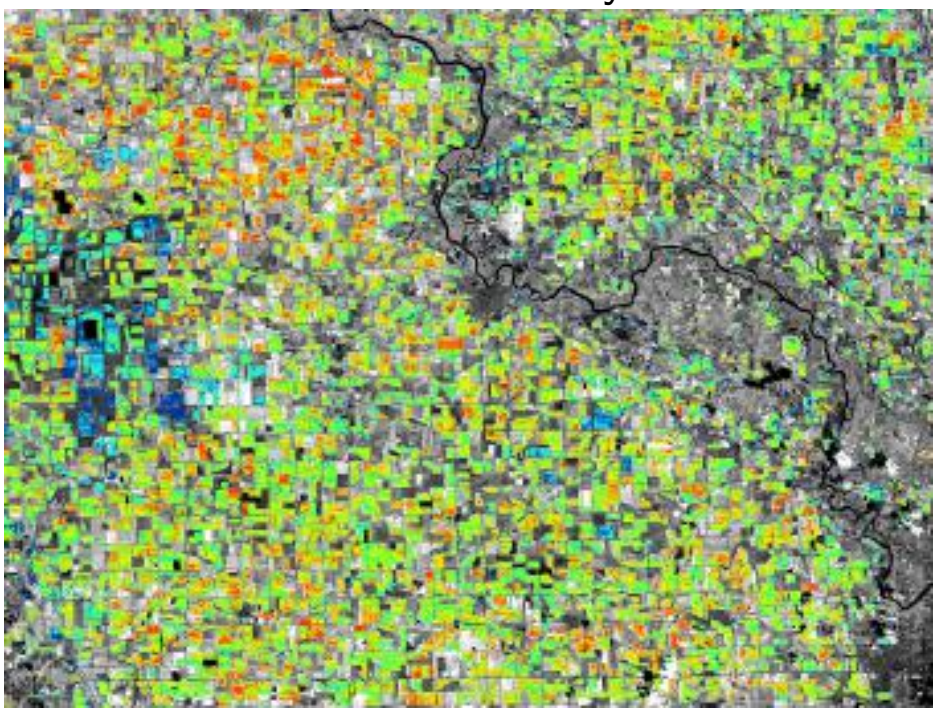



source: Space Angels

# Examples of Business GIS use cases

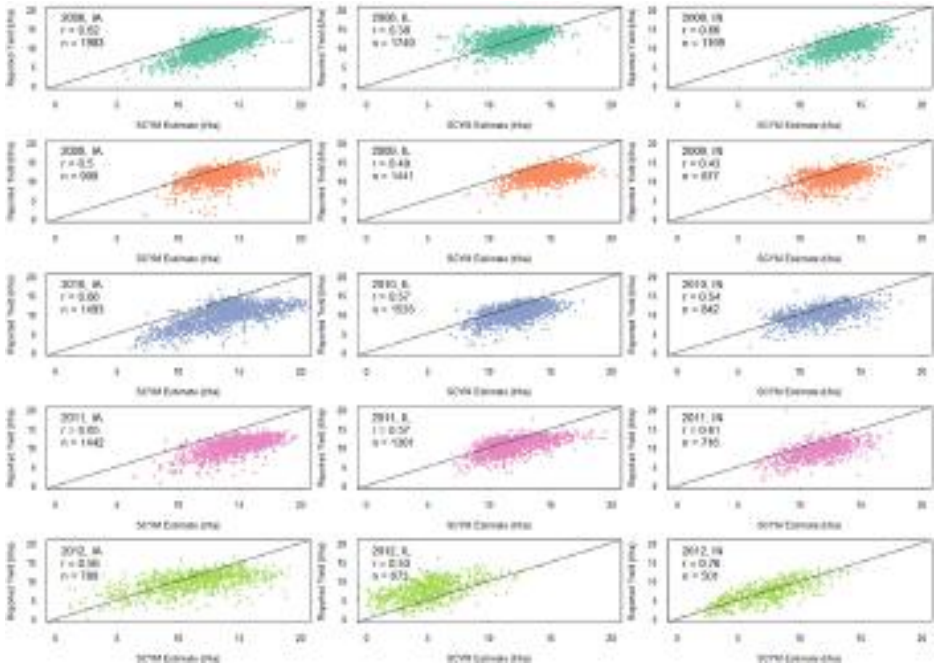
## Farming 1: Yield Measurement

### Satellite estimates of maize yields in Iowa



high yields in red, low yields in blue
source: `http://www.g-feed.com/2015/05/`

### Reported vs. satellite estimates



reported maize yields on the vertical axes, satellite estimates on the horizontal axes; columns: Iowa, Illinois, Indiana, rows: 2008-201

# Examples of Business GIS use cases

## Farming 1: Yield Measurement

**Business:** Corn Farmer

**Goal:** Find out what this season's crop is worth

- in 2016, U.S. produced 13.6 billion bushels of corn on 80.7 million hectares
- corn is used in animal feed, ethanol, corn syrup,...
- starting in August, USDA releases forecasts of corn production
  - monthly frequency
  - state and country-levels
  - surveys farmers, tours fields and creates a forecast

**How GIS can help:** Satellite images + machine learning improve on surveys

- weekly frequency
- geographic granularity (elevators price grain locally)
- survey every field, every day $\rightarrow$ better accuracy

**Examples of firms:** *Descartes Labs* (satellite data provider)

**Other industries:** logistics firms can use better forecasts to plan vehicle fleets, insurers can anticipate risks, commodity traders can forecast prices...

# Examples of Business GIS use cases

## Finance 1: Oil supply

Floating roof oil tanks in satellite imagery

source: Orbital Insight

Shadows identify height of top (oil in tank)

# Examples of Business GIS use cases

## Finance 1: Oil supply

**Business:** Quant fund

**Goal:** Get timely estimates of oil supplies

- Global oil inventories are held in many locations around the world
- Officially reported estimates of oil supplies are problematic
  - Lagged, up to 60 days
  - Biased (incentives to over/under report)
  - Patchy: some regions / countries do not always report

**How GIS can help:** Satellite images + machine learning

- Identify location of floating roof oil tanks
- Height of shadow identifies height of roof $\rightarrow$ amount of oil in tank
- Global coverage, daily frequency, unbiased methodology
- Synthetic Aperture Radar (SAR) to complement electro-optical (EO): see through clouds, at night

**Examples of firms:** *Orbital Insight* (satellite data provider)

# Examples of Business GIS use cases

## Finance 2: Counting objects

Cars in parking lots

source: Digital Globe (`https://bit.ly/2V8DYeQ`)

Ships in ports

Airplanes in airports

# Examples of Business GIS use cases

## Finance 3: Estimating construction activity

Raw satellite image

Size of shadows identifies construction activity

source: Orbital Insight

# Introduction to GIS

## Outline

**We will cover three things to start with GIS**

- Types of geographic data
- Coordinate systems and projections
- Introduction to QGIS

**Geographic data comes in a vast number of formats.**

For 99.9% of all applications you will only need two:

- **Feature** (vector) data, files end in *.shp* (shapefiles)
- **Raster** (cell) data, files (typically) end in *.tif*, but other formats are also common.

It is useful to distinguish three types of feature data

- polygon features
- polyline features
- point features

# Introduction to GIS

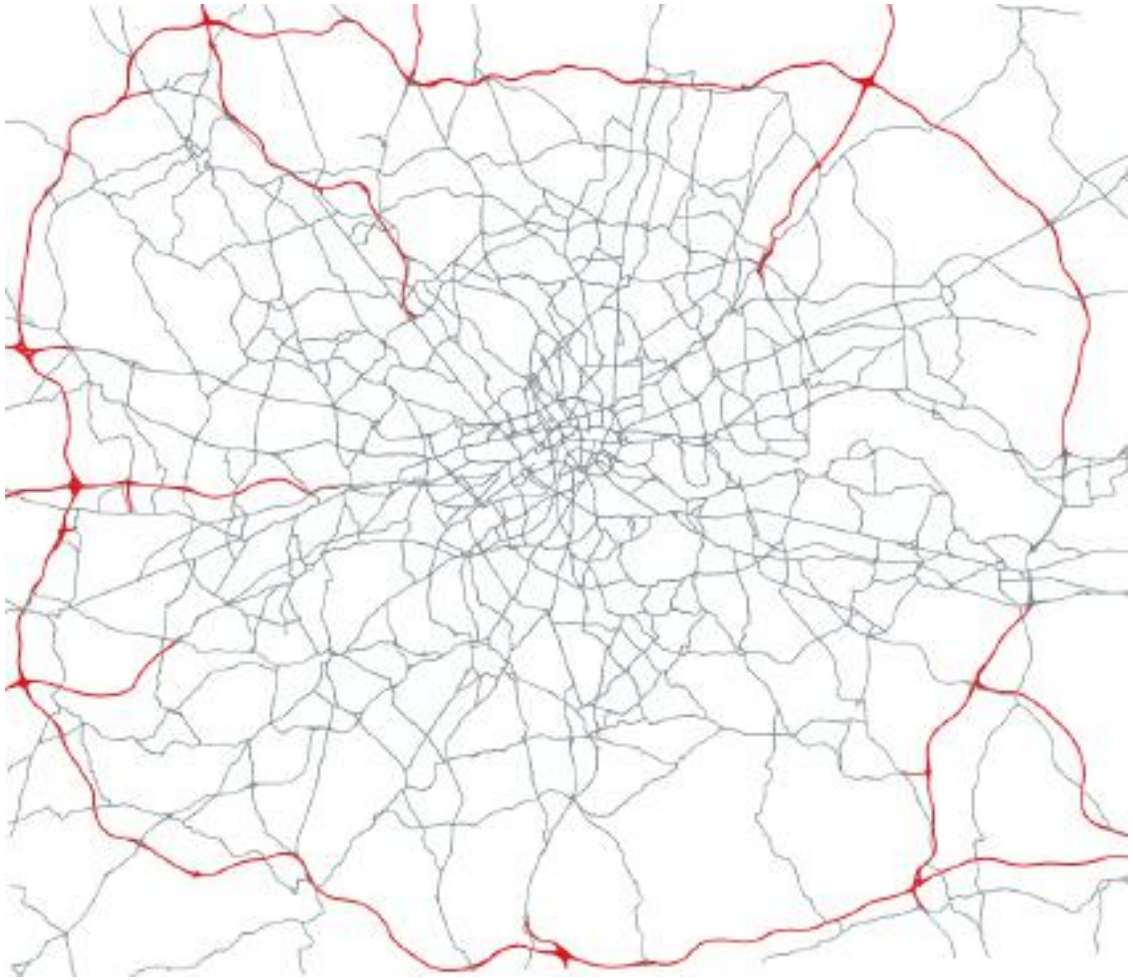
## Data types – polygon features

London's Boroughs can be treated as polygon features

# Introduction to GIS

## Data types – polyline features

Roads can be treated as polylines



# Introduction to GIS

## Data types – point features

Tube stops as points

# Introduction to GIS

## Data types – raster data

$NO_2$ pollution in central London in raster format

Each cell in the raster has a value

| | | | |
|---|---|---|---|
| 1 | 1 | 0 | 0 |
| | 1 | 2 | 2 |
| 4 | 0 | 0 | 2 |
| 4 | 0 | 1 | 1 |

- cells can be empty
- data does not need to be integer (can be float)

**Why this matters**

- Geospatial data are data with geographic identifiers attached to each data point.
- To view, manipulate, and make calculations with geographic data, the identifiers need to be referenced with respect to a coordinate system.
- Each coordinate system represents the Earth's sphere (or a part of it) in two dimensions.
- Results of calculations depend on the coordinate system used. Using the correct one is crucial!

**There are two basic types of coordinate systems**

- Geographic coordinate systems.
- Projected coordinate systems.

Based on Melissa Dell's GIS notes.

Geographic coordinate systems represent locations in spherical coordinates .

It is standard to write information for geographic coordinates systems in two ways

- **degrees, minutes, seconds**
- e.g. London is 51° 30' 26" N, 0° 7' 39" W
- **decimal degrees**
- London is 51.5072, -0.1275
- 30 arc minutes and 26 arc seconds $= 30 \times 60 + 26 = 1826$ arc seconds $= \frac{1826}{3600} = 0.5072$ decimal degrees.

The standard geographic coordinate system is WGS 1984 ("World Geodetic System"). It is by far the most commonly used.

For distances between two points along the Earth's surface, can use **geodesic distance formula**: need only $((lat_1, lon_1), (lat_2, lon_2))$.

**Problem**

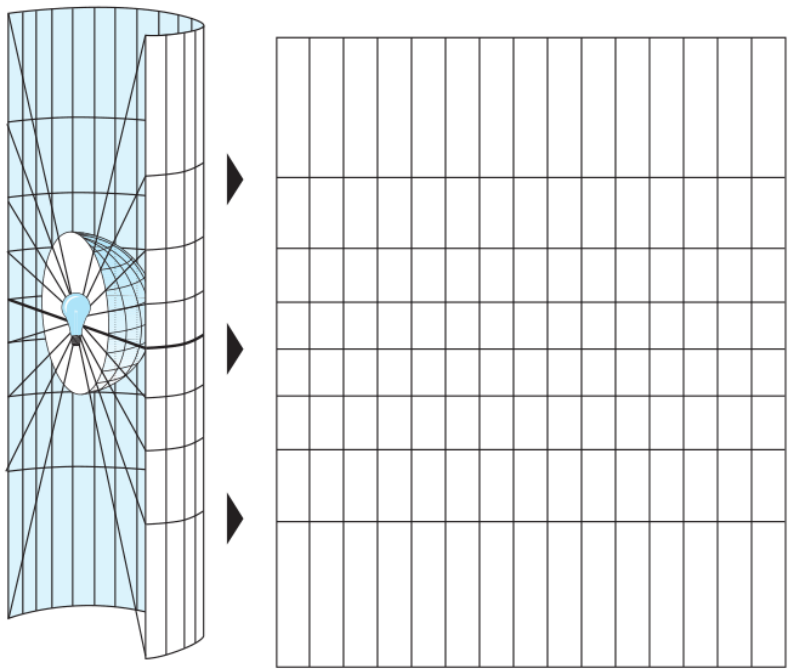
- Geographic coordinate systems are useless to calculate areas and distances along lines.
- Geographic coordinate systems distort lengths and areas: try to wrap a map around an orange!
- $1^\circ$ latitude is 110.6km at the equator, 111.7km at the poles
- $1^\circ$ longitude is 111.3km at the equator, 55.8km at $60^\circ$ N/S

▸ Mercator distortion example

To make accurate area and (some) distance calculations, need to **project** the data from a geographic coordinate system.

One can think of shining a light, placed on the centre of the earth, through the earth surface, and casting a shadow on a projection surface of a certain shape.

Projecting onto a cylindrical projection surface



note stretching of data near the poles

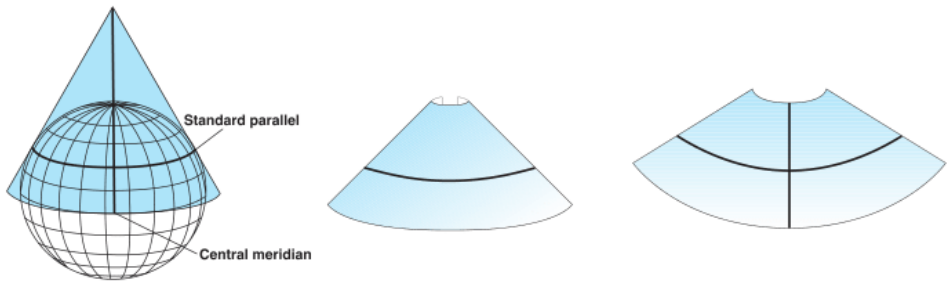
Section based on Melita Kennedy's notes on understanding projections

Useful three-way classification of simple projections

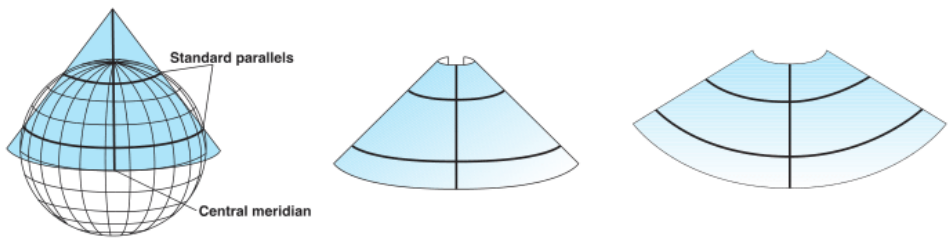
- conic
- cylindrical
- planar

**Conic projections**

Tangent conic projection, one standard parallel

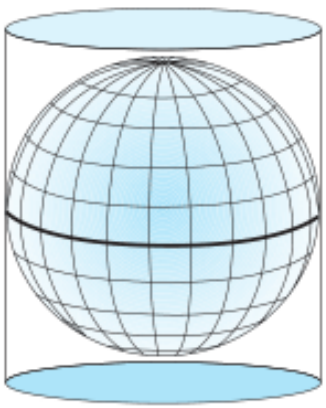



Secant conic projection, two standard parallels



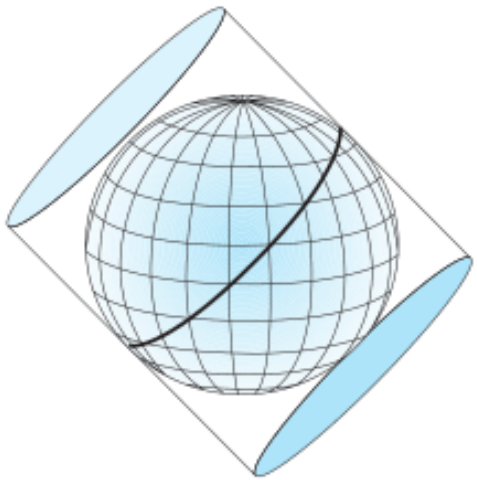
## Cylindrical projections

Cylindrical projections with different lines of tangency



Normal

Transverse

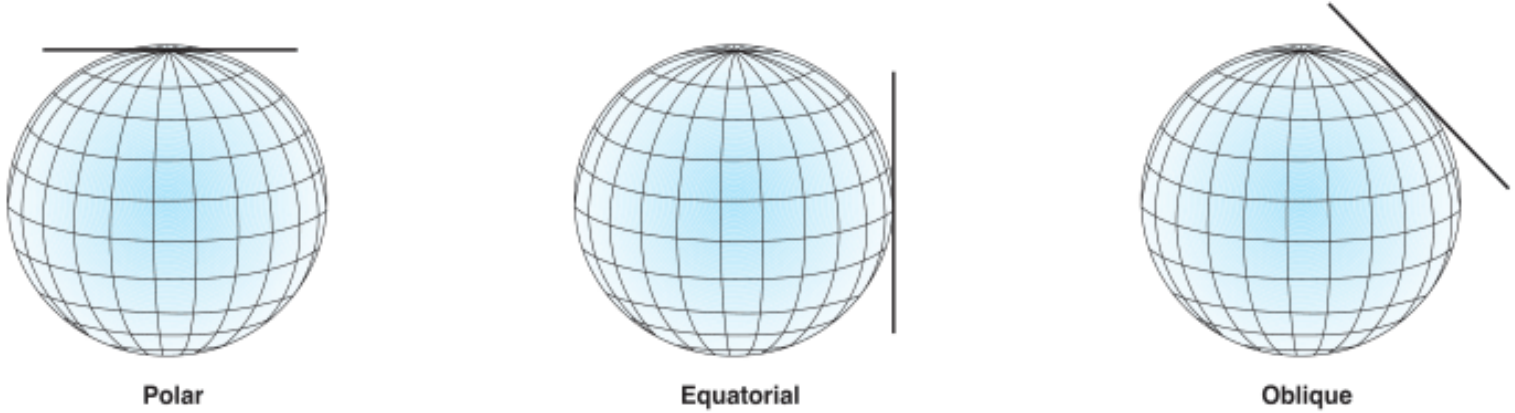


Oblique

Mercator projection is cylindrical with equator as line of tangency.

## Planar projections

Planar projections with different point of tangency (*aspect*)



WGS 1984

- for distance between two points

UTM: divides surface of the earth into many regions, each gets its own projection

- distance/surface area in small regions
- length of polylines

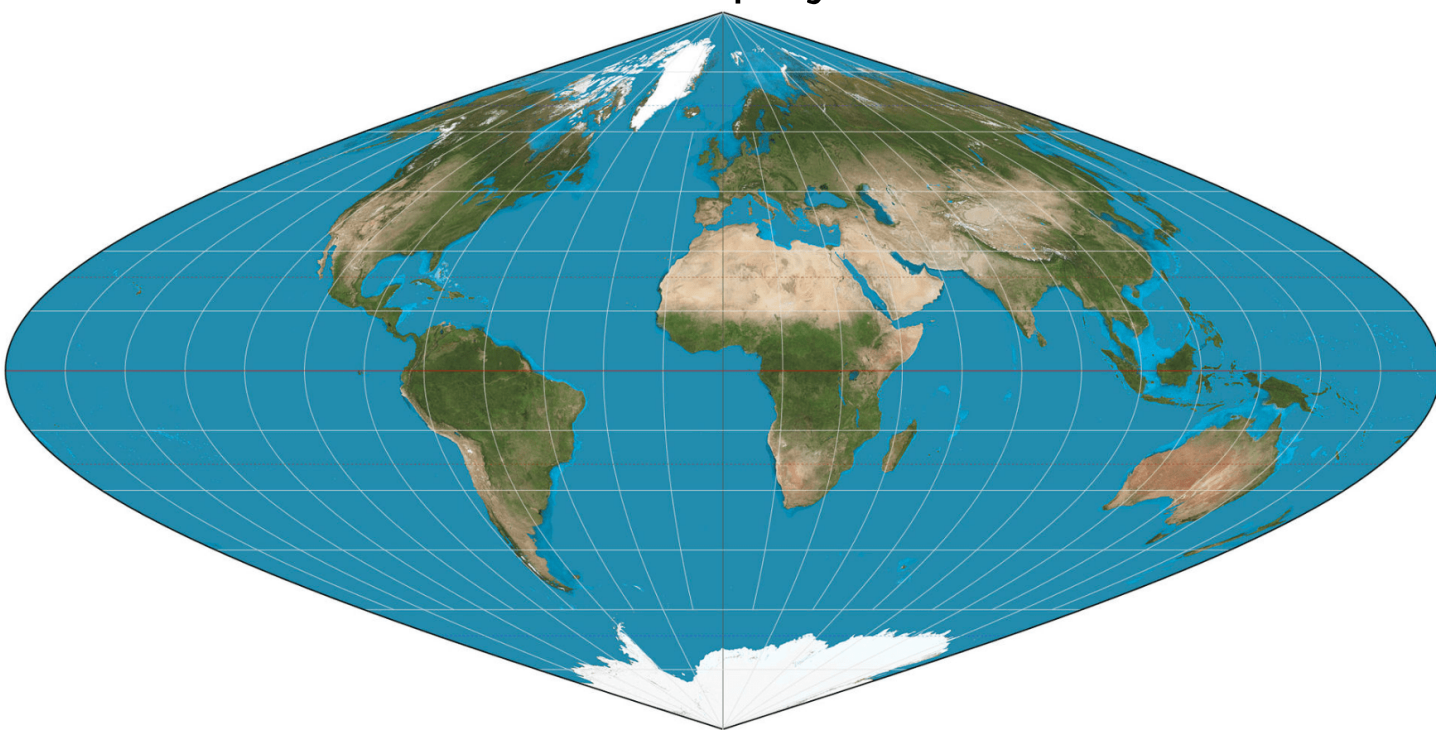
Any equal area projection

- surface area in large regions

This and the following three slides are based on Masayuki Kudamatsu's GIS course

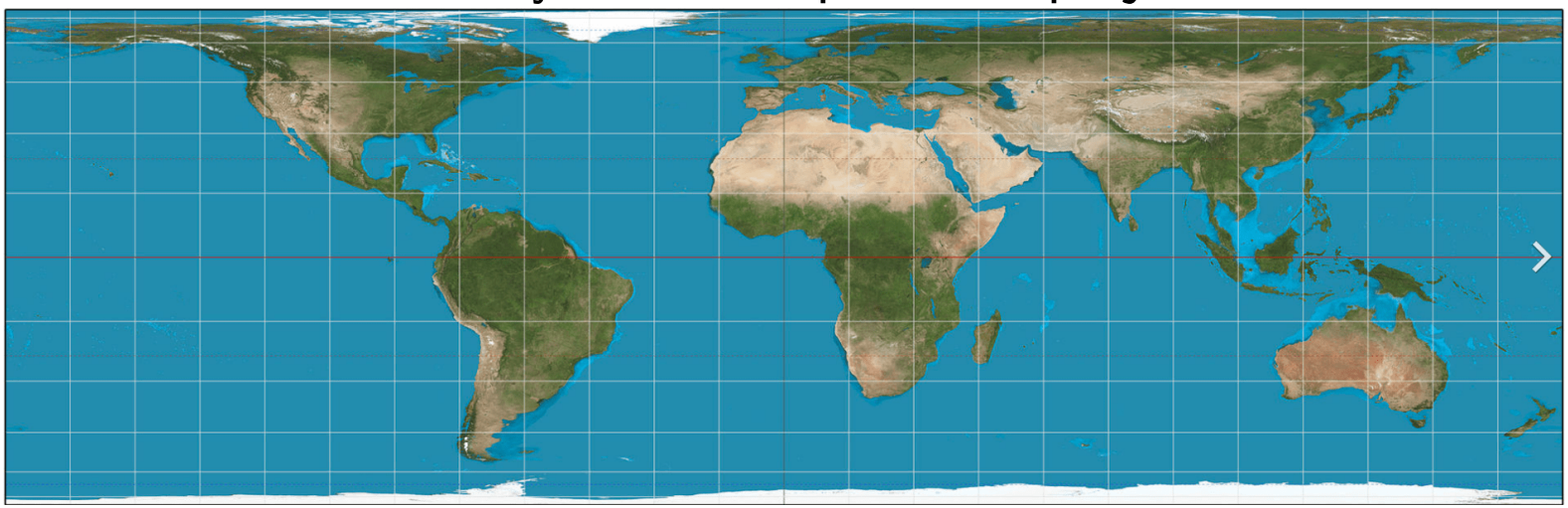
**Differ just in how the world is shown**

Sinusoidal projection



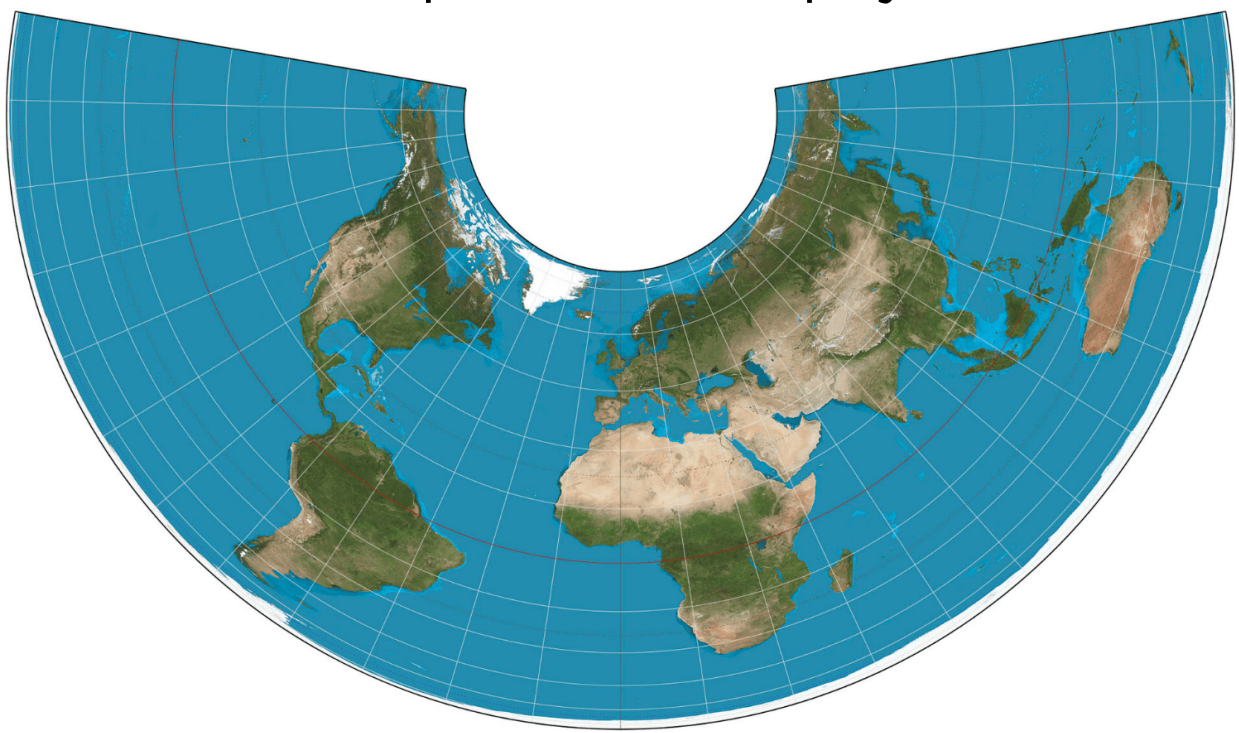
**Differ just in how the world is shown**

Lambert cylindrical equal area projection



**Differ just in how the world is shown**

Albers Equal area conic projection



# Introduction to GIS

## QGIS – What is it good for?

QGIS is a programme for analysing, modifying, and creating geo-spatial data.

Some of the things it can do:

- merge and intersect shapes based on geographical relationships
- finding nearest objects according to specified metric
- calculate slope from elevation data
- take averages within polygons
- finding shortest path in a network
- make maps
- ...

Go here:
`https://qgis.org/en/site/forusers/download.html`

- Go to the tab for your operating system
- Choose the appropriate standalone installer for your system
- We recommend the "Latest release (richest on features):"
- Execute the installer, choose default options, don't install the datasets

**Obstacles for Mac users**

- We have encountered several problems for Mac users
- There is an issue with opening apps from unidentified developers (don't worry, QGIS is safe). See here for a solution `https://kb.wisc.edu/helpdesk/page.php?id=25443`
- There is an issue with `GDAL` tools. See here for a solution `https://gis.stackexchange.com/questions/276853/gdal-scripts-not-found-in-qgis-3-on-osx`
- You may also face an issue with using ORS-tools (see lecture 3). We will address this then.

Intalling QGIS will put a bunch of programmes on your computer.

- We will introduce you to QGIS – the software which allows you to work with geospatial data.
- QGIS has a built in "Browser" – a type of file viewer.
- As you will soon learn, doing anything in QGIS can produce a lot of intermediate files at every step.
- Every shapefile, for example, comes with an *.shp* file (which stores the feature geometry itself), a *.dbf* file (which stores the attribute table, see below), a *.prj* file (which stores the projection), and often with an *.shx* file (which stores a positional index of the feature geometry to allow seeking forwards and backwards quickly).
- The Browser is can make managing this soup of files easier as you always just see one file.

# Introduction to GIS

## QGIS – What the buttons mean


1. "Canvas": data is visualized here
2. "Layers": which datasets are loaded?
3. "Browser": quickly load datasets from disk locations, delete data
4. Processing toolbox

   Note: showing 2-4 is a good default configuration, but other windows can be shown (see "View" → "Panels")

5. Among other things, add layers (data) to canvas
6. Zoom in and out
7. Zoom to full extent (very useful)
8. Get information on geographic features
9. Select elements

# Introduction to GIS

## QGIS – Adding data

**Download some data**

- on google drive
- OR go to `https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london`

  and download "*statistical-gis-boundaries-london.zip (27.34 MB)*", save to some directory and unzip

**Add directory to "Favorites"**

- facilitates loading data
- navigate to the folder containing the file you downloaded and add it to favorites
- sometimes may have to refresh ( ) to see new data

**Add the data**

- In the folder you just added, right-click, double-click (or do the mac thing) to "Add Selected Layer(s) to Canvas" or single click and click on "add layer" ( )
- Can also go more complicated: "Layer" → "Add Layer" → "Add Vector Layer" → "Browser" →navigate to your folder and add the data (ESRI/London_Borough_Excluding_MHW.shp).

# Introduction to GIS

## QGIS – Inspecting the data

You should see a map of London's boroughs in the main window and an entry in "Layers" listing the dataset.
We have loaded a file with **polygon features**. Features come with "attributes".

**Attribute table**

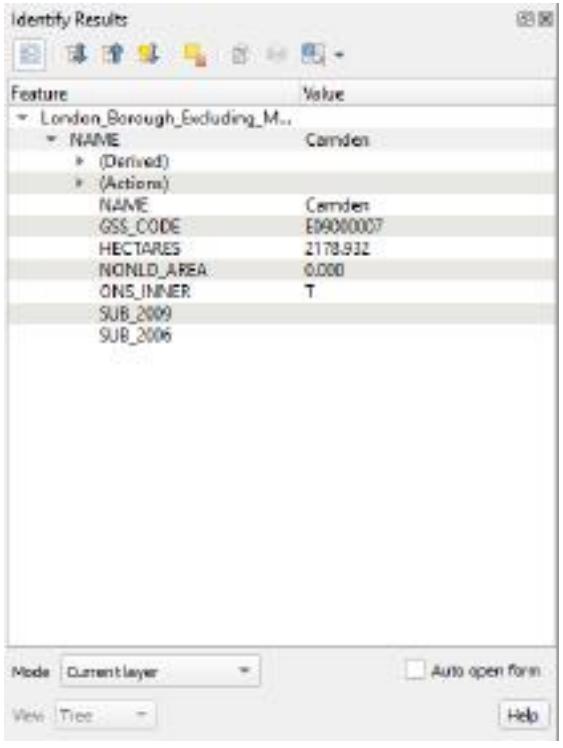
- right click on *London_Borough_Excluding_MHW* in Layers → Open Attribute Table
- a table opens, listing a bunch of variables
- each row in the table corresponds to one polygon feature on the map
- pick a borough and click on the small grey square at the start of its row with the number to select the row and close the attribute table
- the country you picked is highlighted
- click on Zoom to Selection (  )
- click on  to zoom back out
- click on Select Features (  ) control-click on a feature on the canvas to un-select it (or use  )
- play around with zooming, panning (  ), selecting, until you are comfortable

# Introduction to GIS

## QGIS – Inspecting the data

We can also use **Identify Features** to view the information in the attribute table for one or more features

- click on  and click on Camden
- you should see the window on the right
- this lists all the variables in the attribute table for this particular feature
- we can select more than one feature this way (hold down the mouse and drag it over several features)
- then expand and collapse individual feature attribute lists to find the one(s) we are interested in



## Extent and coordinate system

- back in Layers, right click on the layer and "Properties" at the bottom
- select the "Information" tab and check "Information from provider"
- select the "Source" tab


# Introduction to GIS

## QGIS – Changing color, outline width, labelling features

**Change color of boroughs and outline width**

- Back under Layers, right click the *London_Borough_Excluding_MHW* layer → Properties → Symbology
- Change Fill color, outline ("Stroke") color, and outline width
- Click "OK" and "Apply" and see how the canvas display changes

**Label features**

- Back to Layer Properties window, select the "Labels" tab
- From the drop-down menu at the top, choose "Single Labels"
- From the "Label with" drop-down menu, select "NAME" (this is one of the variables from the attribute table)
- Click "OK" and "Apply" and see how the canvas display changes

**Hide a dataset**

- in Layers, uncheck *London_Borough_Excluding_MHW*

# Introduction to GIS

## QGIS – Adding and inspecting some raster data

### Download some data

- go to google drive
- download the data on ground-level concentrations of emissions, save to some directory and unzip
  note: we have transformed these data so you can open them in QGIS: the original data source is `https://bit.ly/2UgQ7mN`

### Add the data for $NO_2$ concentration

- exactly as with the feature data (the data are under *LAEI2016_2016_NO2.tif*)

### Inspect the data

- zoom in closely so you can make out individual cells
- use  to look up individual pixel values
- look in the source tab of the Properties menu

### Change the color scheme

- Layer → Properties → Symbology
- Band Rendering → Render Type → Singleband pseudocolor → choose Linear Interpolation and your favorite Color Ramp → OK + Apply

So far we have added data downloaded from the web. There is one type of spatial data that we can easily create ourselves: **Point features**.

**Create point features in text editor**

- open a text editor
- in the first line, type: *point_name, latitude, longitude*
- in the second line, type: *some name, 51.5042, -0.13357*
- in the third line, type: *some other name, 51.6252, -0.23001*
- keep adding as many points as you like, until you're bored.
- make sure latitude $\in [-90, 90]$, longitude $\in [-180, 180]$
- save the data under *my_points.csv* in the directory where you saved the files you downloaded

# Introduction to GIS

## QGIS – Adding and inspecting latitude/longitude data

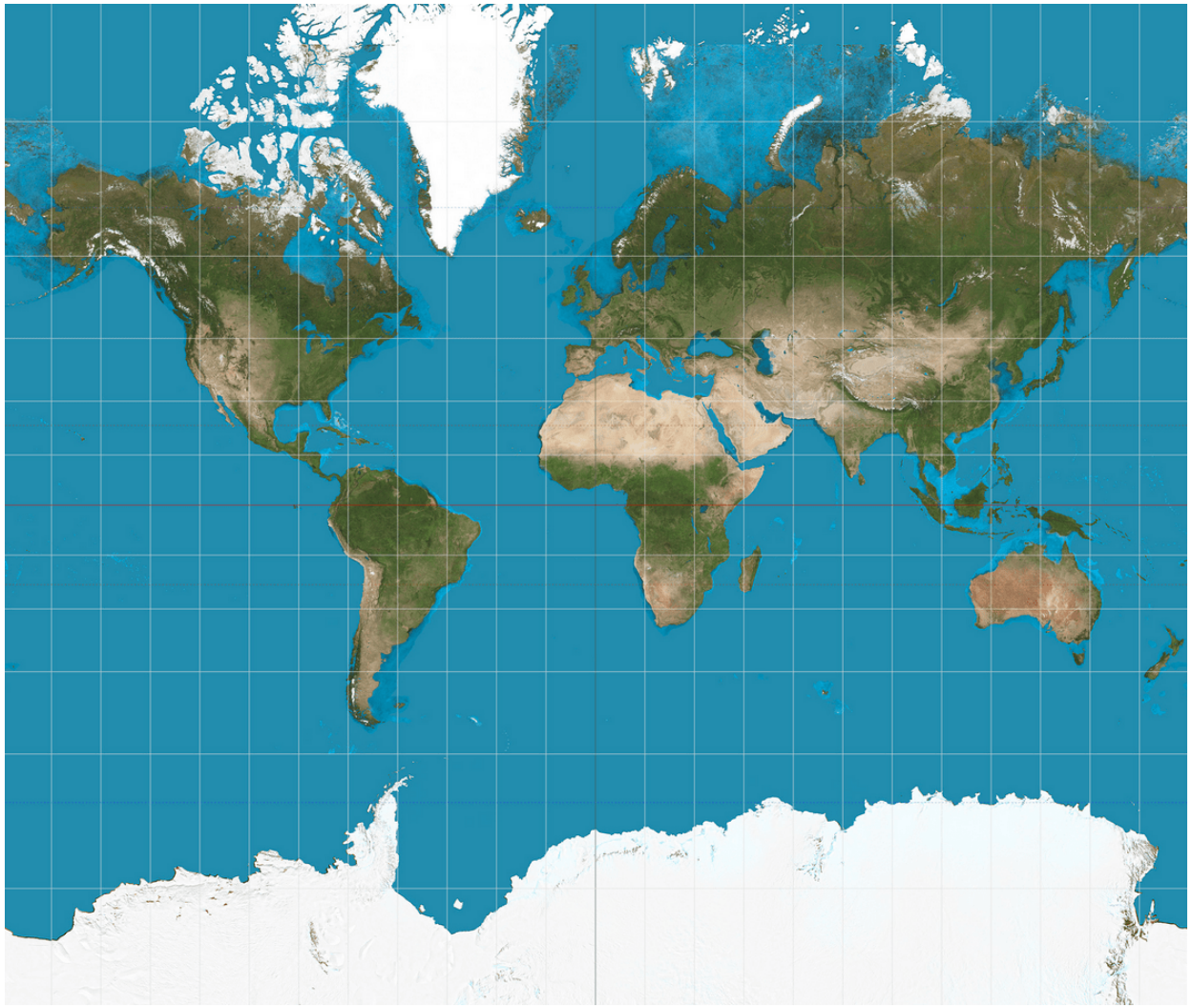
### Add and display data

- Layer → Add Layer → Add Delimited Text Layer → Under File name, browse to the point layer; Pick a layer name
- File Format → CSV
- Geometry Definition → Point coordinates
  - X field: longitude
  - Y field: latitude
  - Geometry CRS: Project CRS: EPSG: 4326 – WGS 84
- Click "Add"
- dots appear on the map

### Inspect the data

- look at the attribute table
- use Identify Features (  ) to get information on individual points

The world in Mercator projection



## The actual sizes of Greenland and Australia


nice *The Economist* article on map projections:

`https://www.economist.com/blogs/graphicdetail/2016/12/daily-chart-1?fsrc=scn/fb/te/bl/ed/misleadingmapsandproblematicprojections`

back