

# EDA and ESDA

Esteban Lopez Ochoa  
Master in Business Analytics

06/04/2020

- Exploratory Data Analysis (EDA)
  - “discover potentially explicable Patterns” (Good, 1983)
  - Data Visualization
    - Tables
    - Summary statistics
    - Charts
    - Maps
    - Dynamic Graphs/Maps
    - Brushing and Linking
  - Data has multivariate nature
    - This problem has increased since the Big Data era.
    - But “looking” at the data can save you many hours of wasteful modeling work
  - “detect the expected and discover the unexpected”
    - Thomas et al (2005)

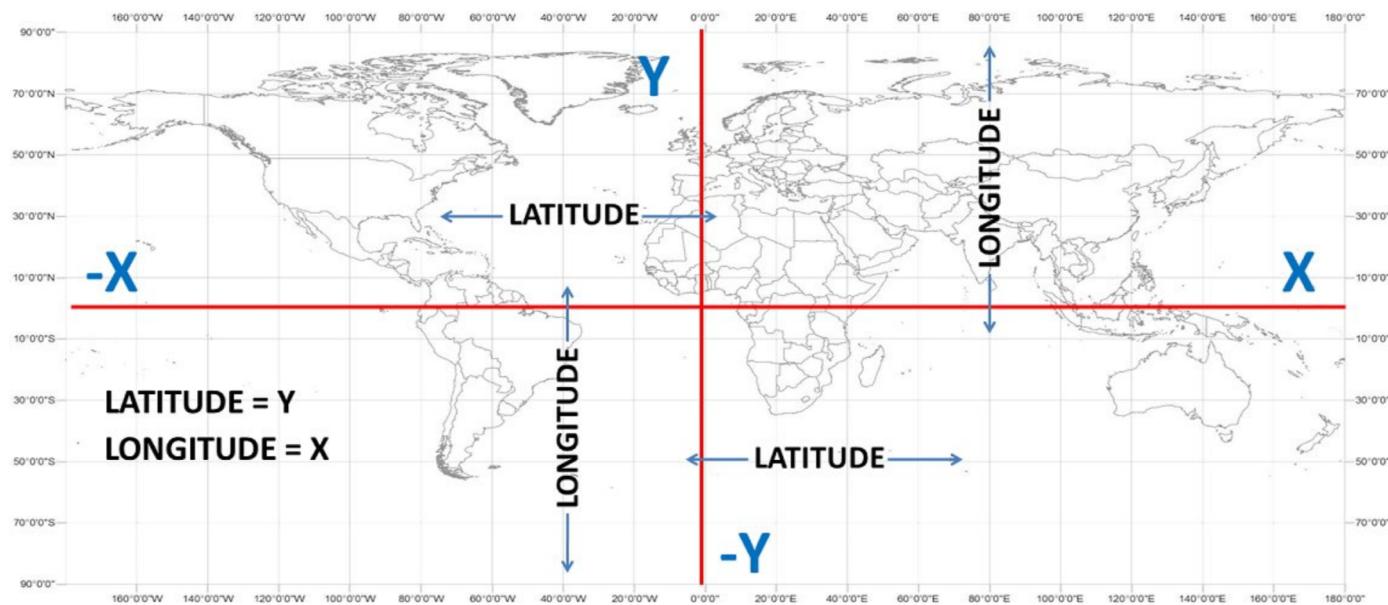
- Exploratory Spatial Data Analysis (ESDA)
  - Extended EDA to space
  - More than just mapping
    - Mapping – take a variable and plot it (results)
  - ESDA Definition
    - Understand that spatial information as a key component of the data exploration process
    - Focuses on spatial patterns
  - ESDA Activities
    - Describe spatial distributions – dynamic maps
    - Identify atypical observations – spatial outliers
    - Discover patterns of spatial dependence and spatial heterogeneity
      - Spatial clusters, hot/cold spots
      - Spatial structured breaks
      - Spatial aggregation – Functional Labor Market areas

- Scatter Plot – Brushing and linking
  - <https://shiny.rstudio.com/articles/plot-interaction.html>
- 3D plot
  - <https://testing-apps.shinyapps.io/3devents/>
- Parallel Coordinate Plot
  - <https://testing-apps.shinyapps.io/shiny-parcoords/>
  - <https://www.r-graph-gallery.com/parallel-plot.html>
- Scatter plot matrix
  - <https://testing-apps.shinyapps.io/shiny-corrplot/>
- Conditional Plots
  - facet\_wraps in ggplot

# MAPPING BASICS

# Maps files

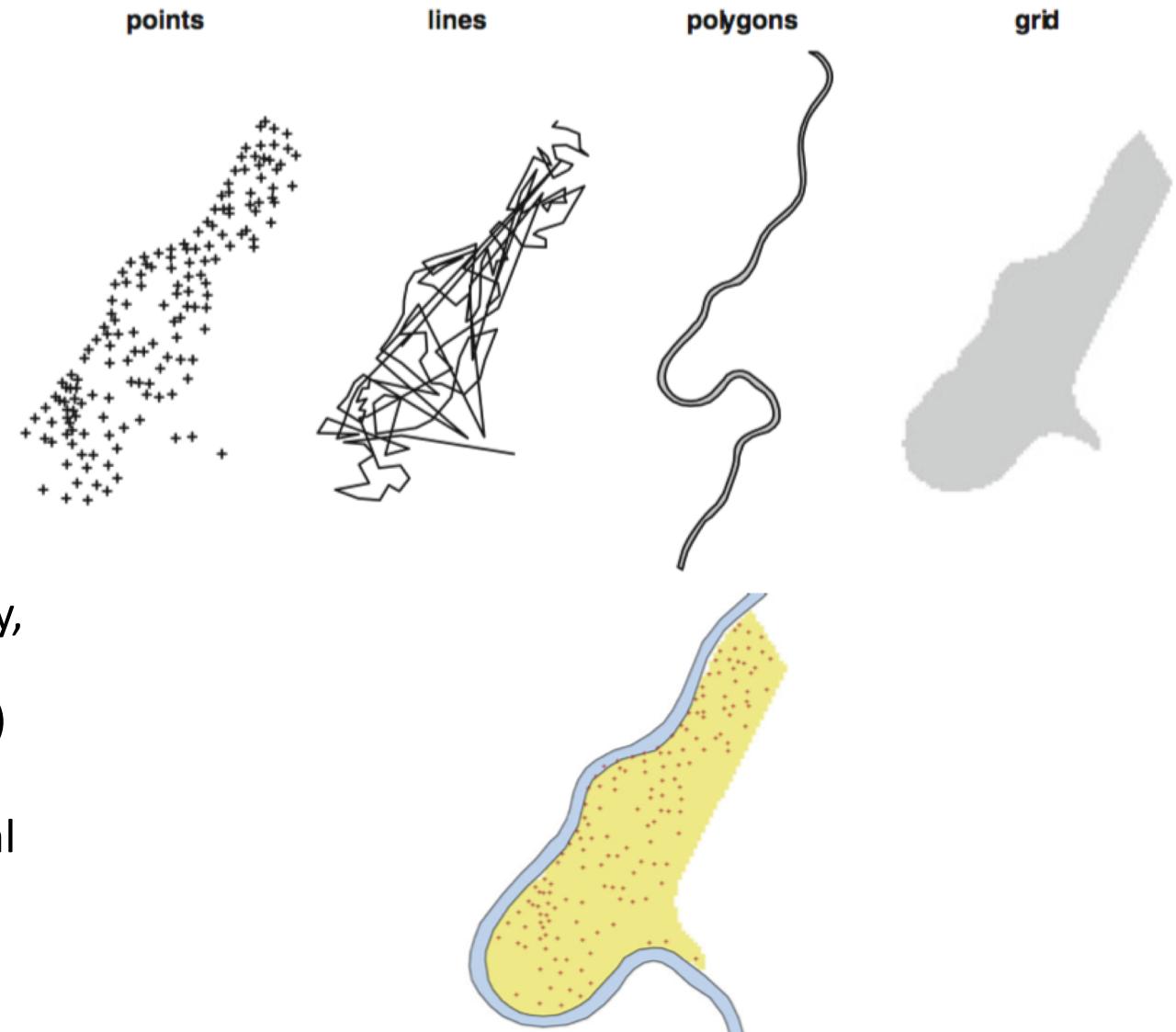
- Maps are representations of multivariate objects
  - Location: Longitude (X) & Latitude (Y)
  - Shape/feature: points, polygons, surfaces, lines
  - Data: each entry for each feature
  - Projection: Geographical or Projected
  - Units: Meters, kilometers, feet



# Types and Sources

- **Types**

- Points
- Lines
- Polygons
- Surfaces  
(raster)
  - Satellite –  
(e.g. imagery,  
weather,  
oceans, etc.)
  - Topographic  
– (e.g. Digital  
Elevation  
Models -  
DEMs)

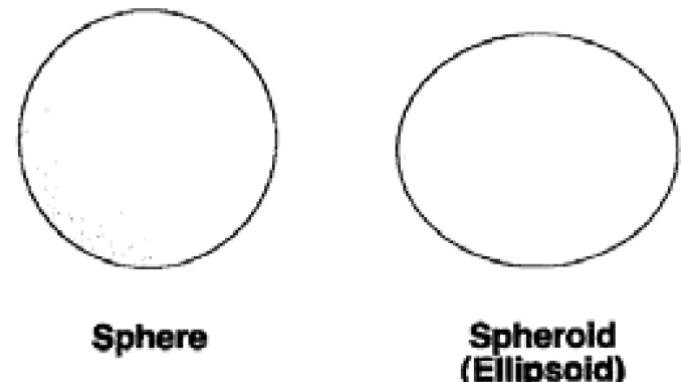


# Types and Sources

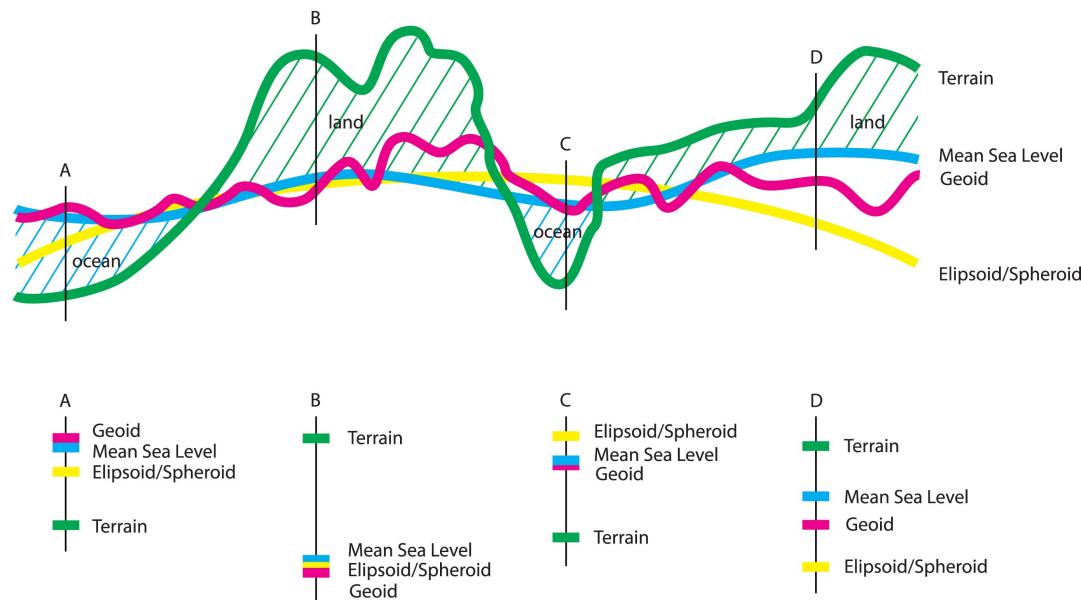
- Sources
  - Points, Lines, Polygons (of Chile)
    - General: [www.ide.cl](http://www.ide.cl); <http://www.geoportal.cl/visorgeoportal/>
    - CONAF: <https://sit.conaf.cl>
    - <https://www.sitantofagasta.cl>
  - Surfaces (raster) (of Chile)
    - SAF: Servicio Aéreo Fotogramétrico de Chile (<https://www.saf.cl>)
    - IGM: Instituto Geográfico Militar (<https://www.igm.cl>)
    - USGS: United States Geological Survey - NASA  
(<https://earthexplorer.usgs.gov>)
    - ESA: Sentinels Satellites - Copernicus Mission  
<https://sentinels.copernicus.eu/>

# Coordinate Systems & Map Projections

- Geodesy = dividing the earth
  - “The science that determines the figure of the earth and the interrelation of selected points on (or near) its surface” (Smith, 1997)



- Geoid
  - Most faithful aprox of Earth's shape



- Ellipsoid
  - Mathematical model of earth's surface, a geoid's simplification
- Datum
  - Horizontal system for accurately measuring distances on earth's surface
  - Most common: WGS84, NAD27
  - Changes to the values of datum parameters result in changes to coordinate values of points
  - Satellite technology has made possible to have earth-centered datums (e.g. NAD 83)

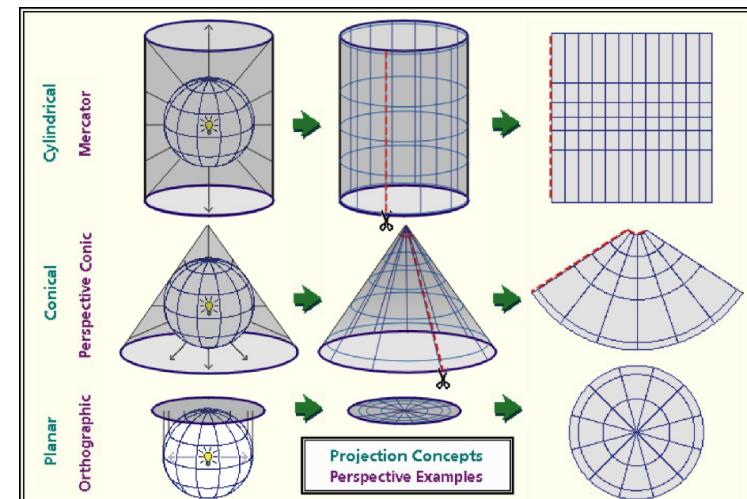
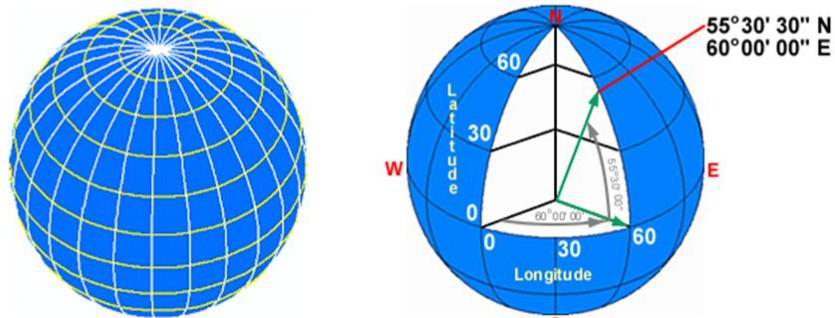
- Coordinate Systems

- Two-dimensional representation of a three-dimensional space
- Specifies:

- Origin
- Orientation
- Units of Measure

- Types:

- Geographic – latitude and longitude to define location of points in the surface of a 3D spheroid (has: datum, units, prime meridian)
- Projected – Flat surface (2D) that has been projected to represent equal lengths between angles

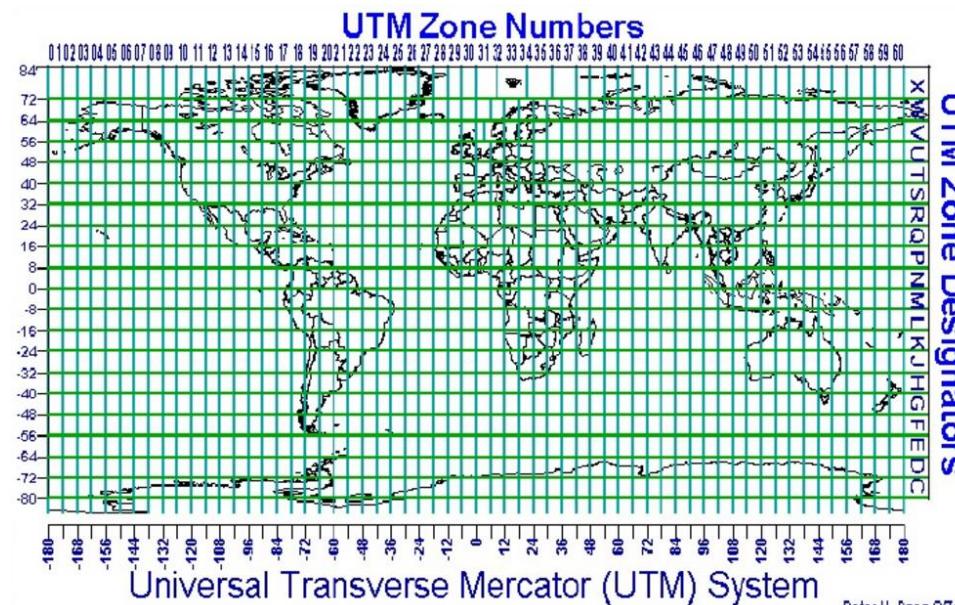


- **Projections:**
  - Equal-area: preserve areal relationships
    - Thematic or distribution maps
  - Conformal: preserve shapes, directional relationships
    - Presentation Maps
  - Equidistant: preserve distance relations
    - Navigational maps (mercator)

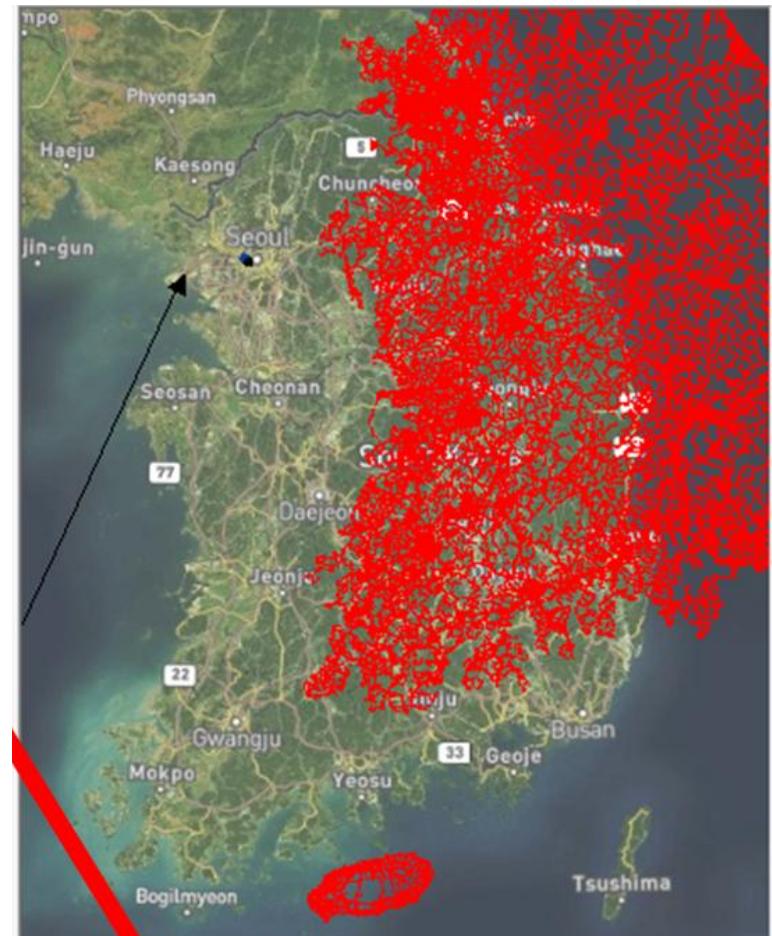
## • Scale of coordinate systems

Global: UTM – Universal Transversal  
Mercator

Divide world in 6 strips each  
spanning 6 degrees of longitude  
False eastings and northings make  
all coordinate positive

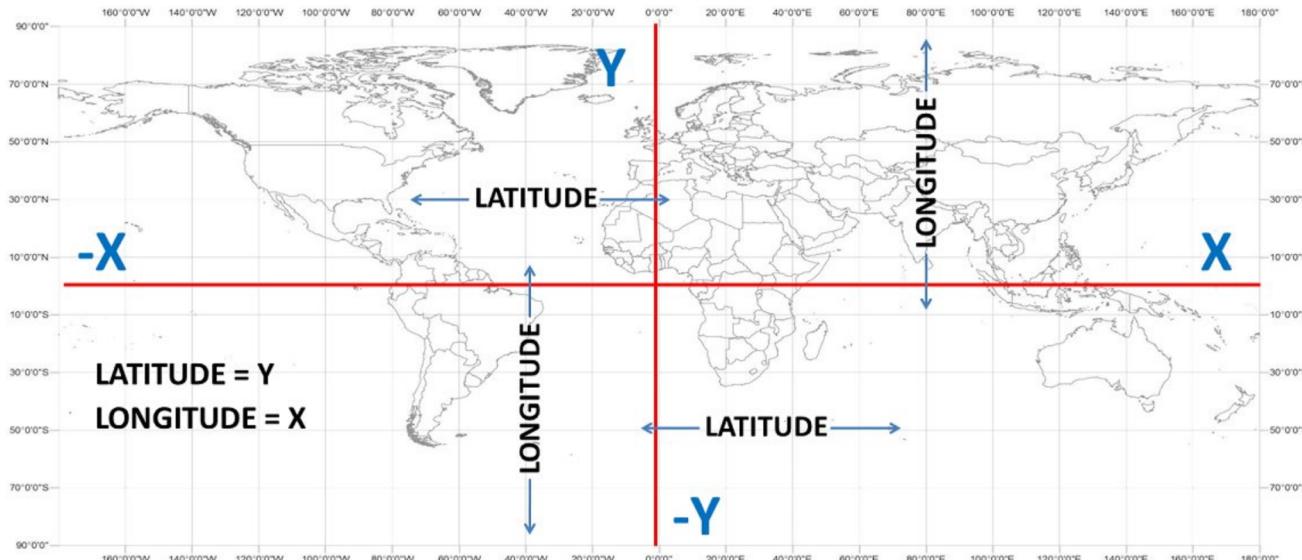


- Mapping data with different projections
  - Some data can be referring to the same place but having different coordinate systems and different projections
  - Overlaying data requires a single projection
  - Transforming one coordinate system to another is a must for geographic overlay and distance calculations

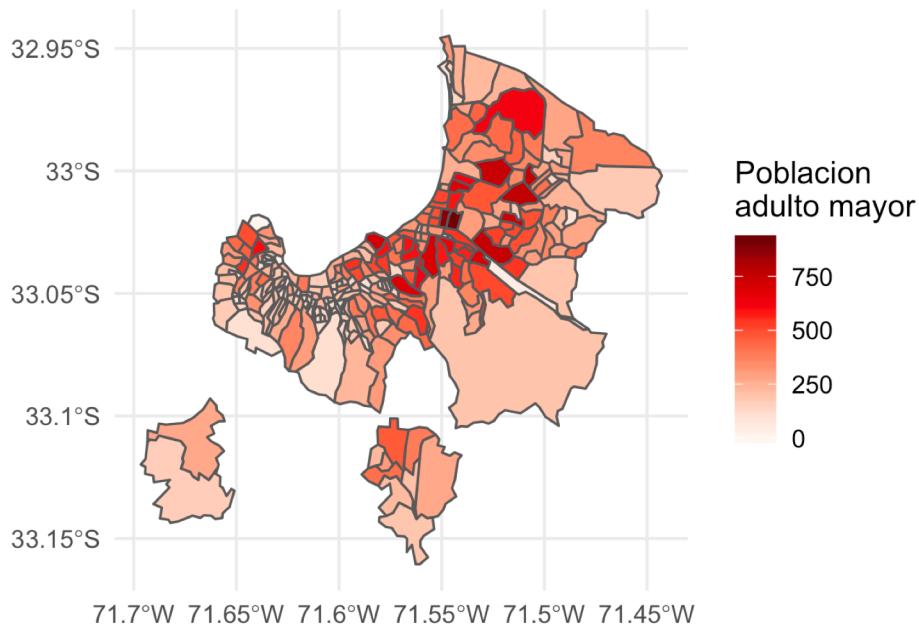


# Maps files

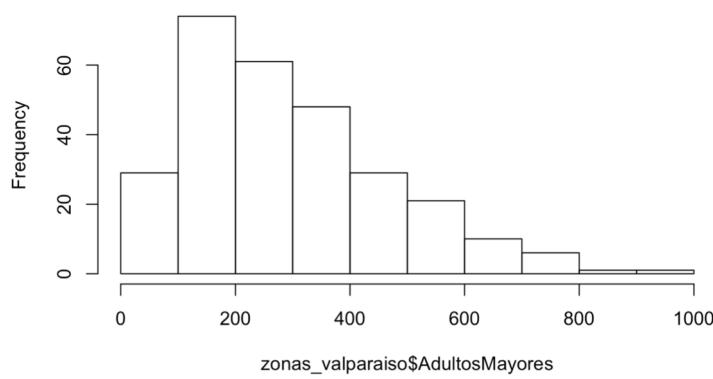
- Maps are representations of multivariate objects
  - Location: Longitude (X) & Latitude (Y)
  - Shape/feature: points, polygons, surfaces, lines
  - Data: each entry for each feature
  - Projection: Geographical or Projected
  - Units: Meters, kilometers, feet
  - Format or File extension: \*.shp, \*.kml, \*.kmz,
  - Manipulation format: SpatialDataFrame (sp) vs. SimpleFeatures (sf)



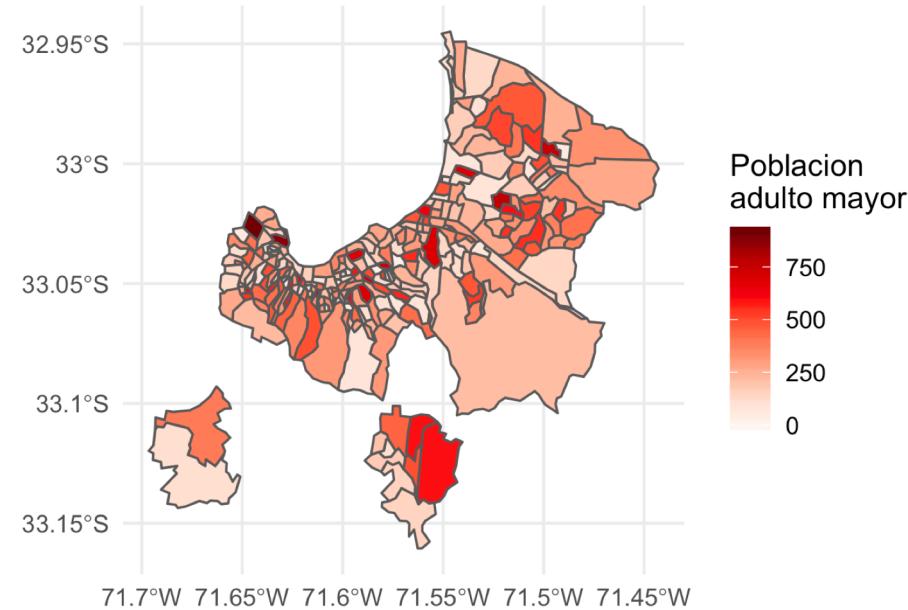
Poblacion de 65 años y más  
Valparaíso y Viña del Mar



Histograma Adultos Mayores Viña-Valpo



Poblacion de 65 años y más  
Valparaíso y Viña del Mar



Histograma Adultos Mayores Viña-Valpo

