

EDA and ESDA

Part 2

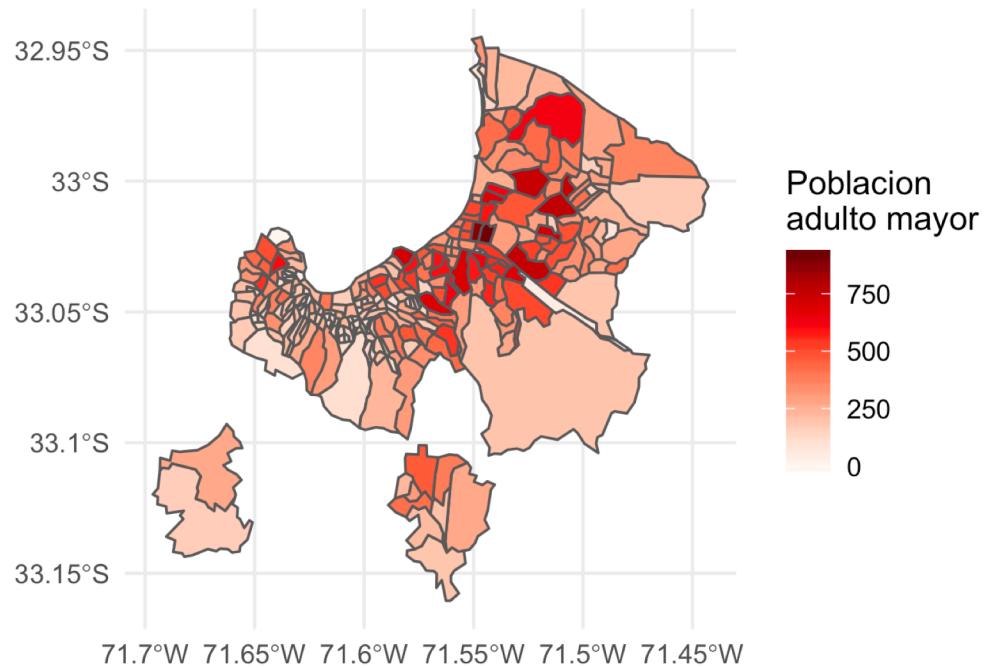
Esteban Lopez Ochoa
Master in Business Analytics

12/04/2020

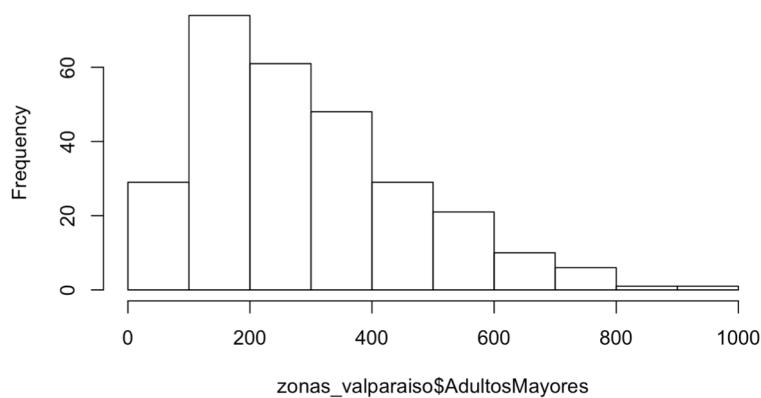
- Exploratory Data Analysis (EDA)
 - “discover potentially explicable Patterns” (Good, 1983)
 - Data Visualization
 - Tables
 - Summary statistics
 - Charts
 - Maps
 - Dynamic Graphs/Maps
 - Brushing and Linking
 - Data has multivariate nature
 - This problem has increased since the Big Data era.
 - 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

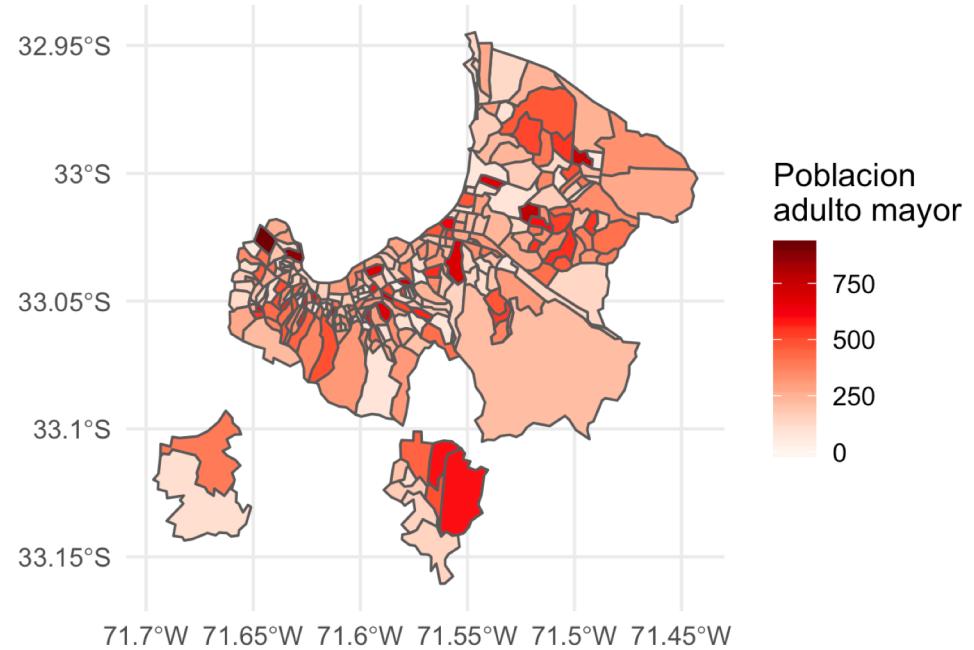
Poblacion de 65 años y más
Valparaíso y Viña del Mar



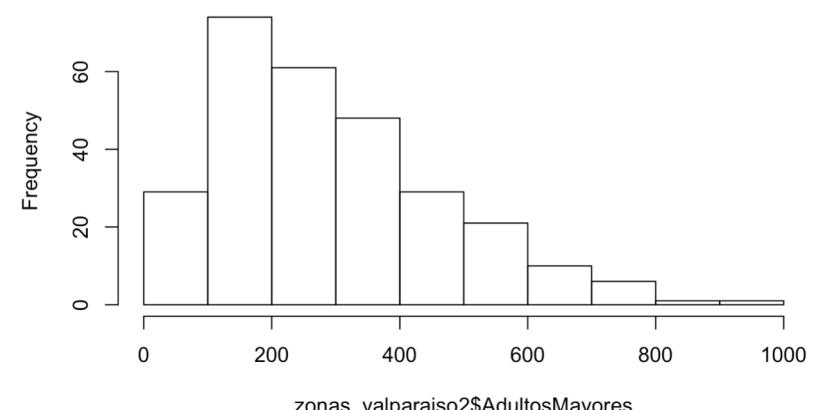
Histograma Adultos Mayores Viña-Valpo



Poblacion de 65 años y más
Valparaíso y Viña del Mar



Histograma Adultos Mayores Viña-Valpo



ESDA: Describe spatial distributions

- Map classifications

- Defined Intervals

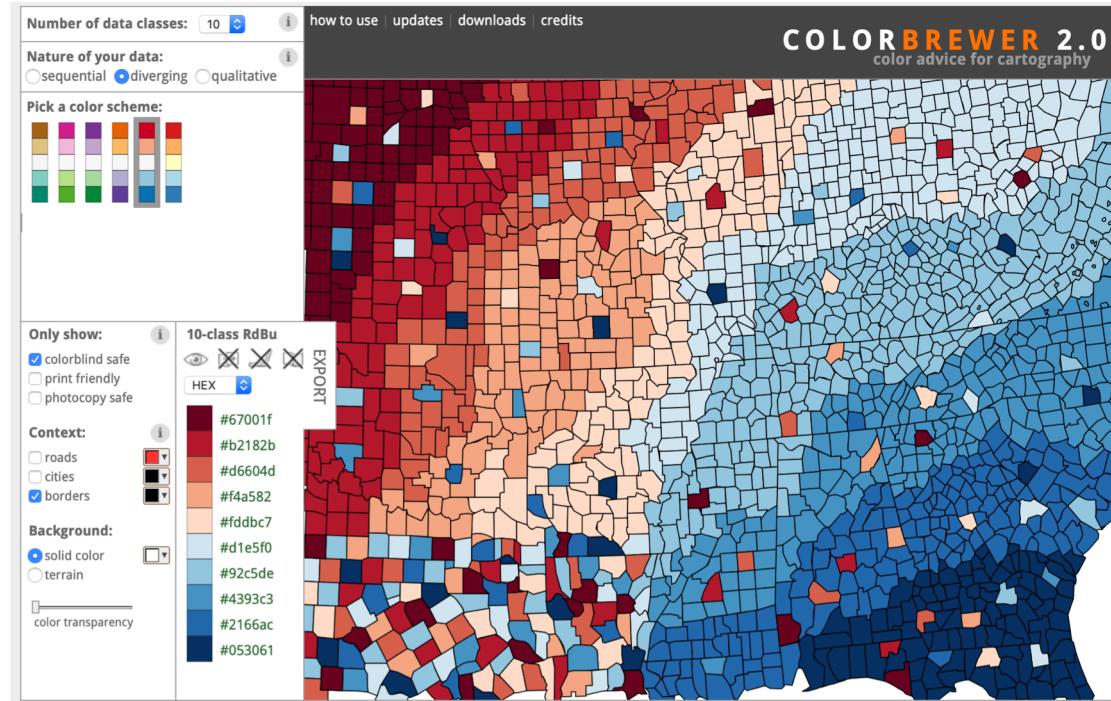
- Manual
 - Equal interval
 - Natural Breaks (Jenks)

- Statistical based Intervals

- Quantile (equal share)
 - Standard Deviation
 - Extreme values

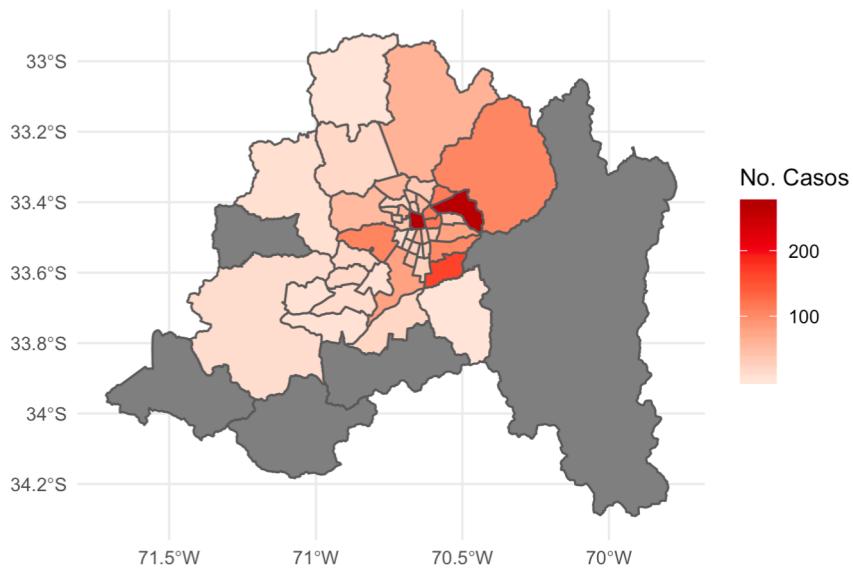
- Keep in mind when mapping

- Color of choice (colors have meanings!)
 - Projection of choice (distance vs. shape based representations)
 - Definition of categories (non-ordered, ordinal, divergent, etc.)



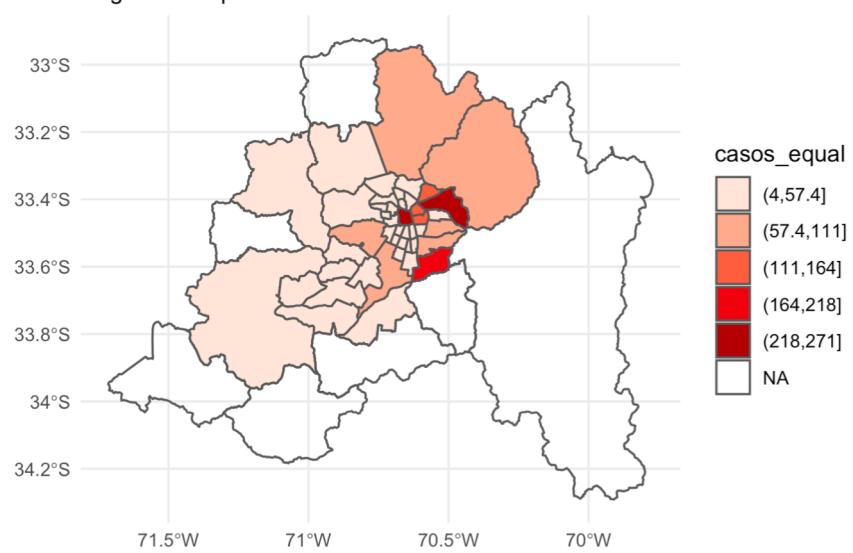
Casos Confirmados

Región Metropolitana - 2020-04-08



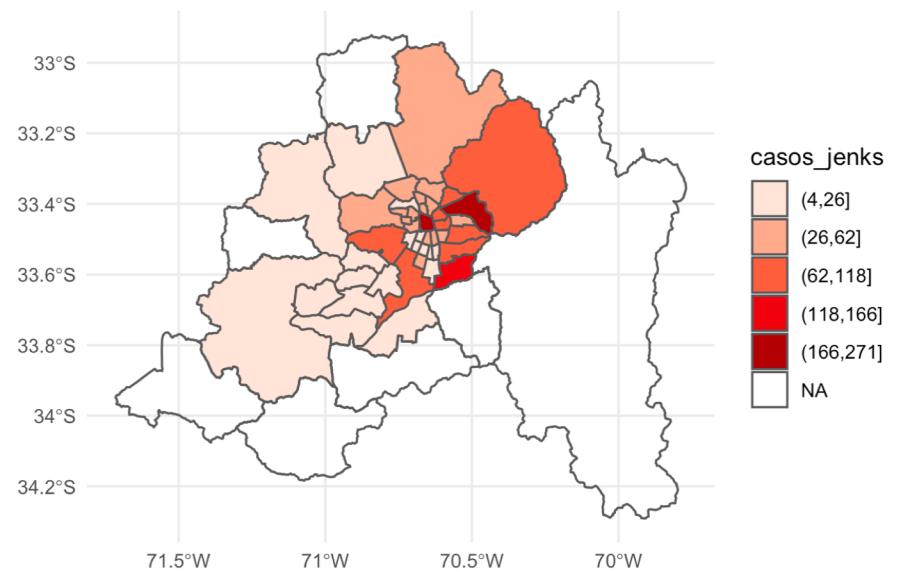
Casos Confirmados

Región Metropolitana - 2020-04-08



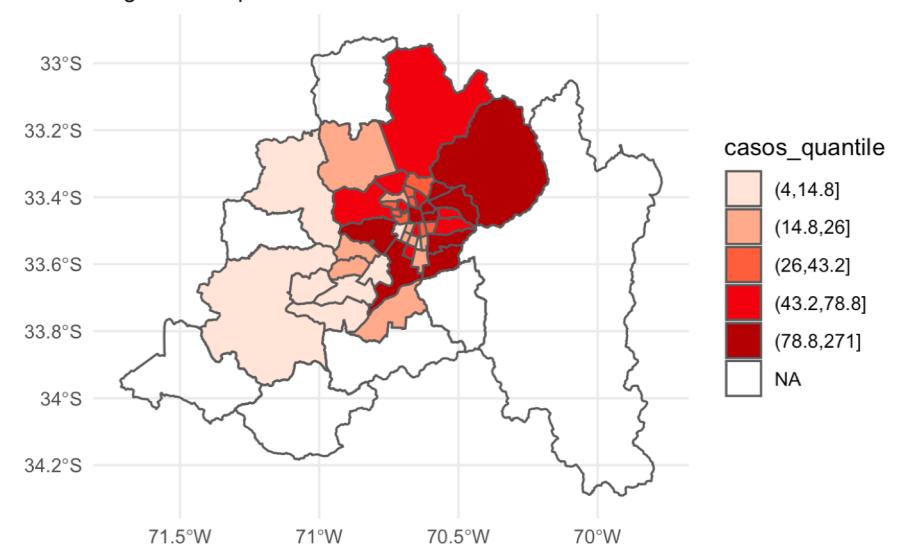
Casos Confirmados

Región Metropolitana - 2020-04-08



Casos Confirmados

Región Metropolitana - 2020-04-08



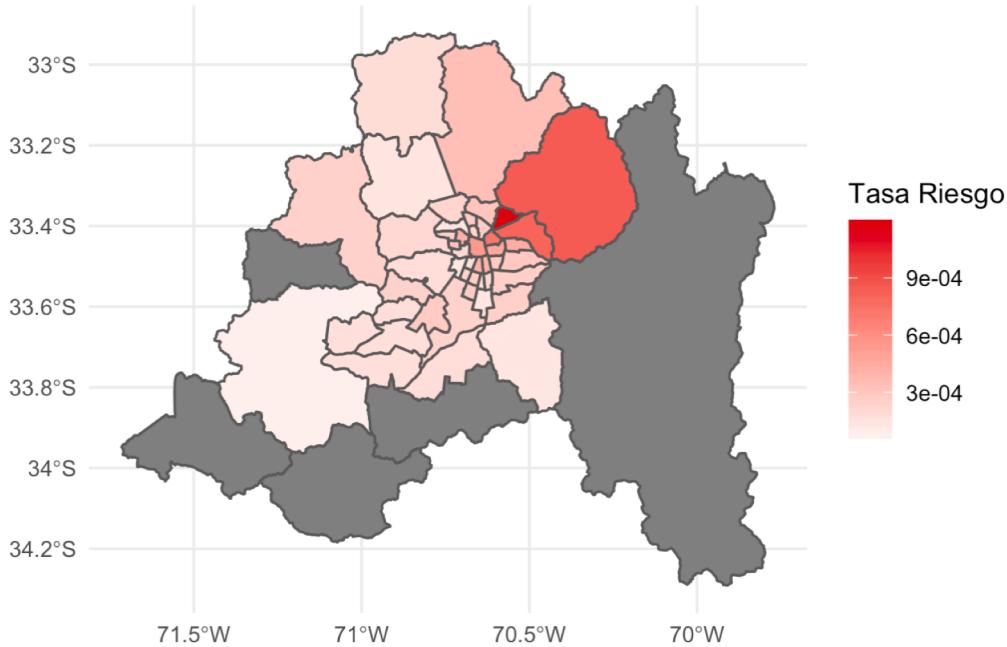
ESDA: Describe spatial distributions – Mapping rates

- Riesgo y tasas
 - Riesgo: probabilidad de que un evento ocurra
 - El riesgo real no se observa
 - Solo se observa eventos
- Estimación del Riesgo
 - Tasa cruda (raw rate)
 - O_i : No. de Eventos
 - P_i : Población en riesgo
 - r_i : O_i/P_i
 - Foco → Heterogeneidad Espacial
 - El riesgo es uniforme a través del espacio?
 - Dónde hay riesgos elevados?
 - Qué causa esas altas tasas de riesgo?

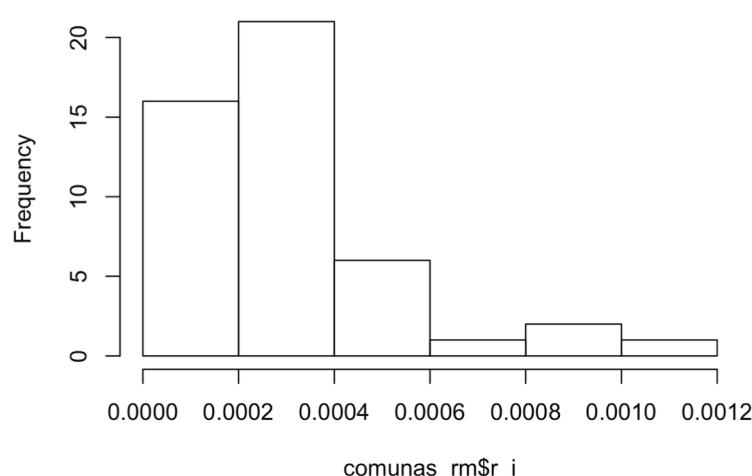
ESDA: Describe spatial distributions – Mapping rates

- Riesgo promedio (AR)
 - $AR = \sum(O_i) / \sum(P_i)$
- Eventos esperados
 - $E_i = AR * P_i$
- Riesgo Relativo
 - $R_i = O_i / E_i$
- Exceso de Riesgo
 - Idea básica: riesgo mayor que algún estandar
 - Estandar: algun grupo de referencia
 - e.g: Población de la comuna, Nicho de Mercado, etc.
 - $R_i > 1 \rightarrow$ Exceso de Riesgo (más eventos que el promedio)
 - $R_i < 1 \rightarrow$ Bajo Riesgo (menos eventos que el promedio)

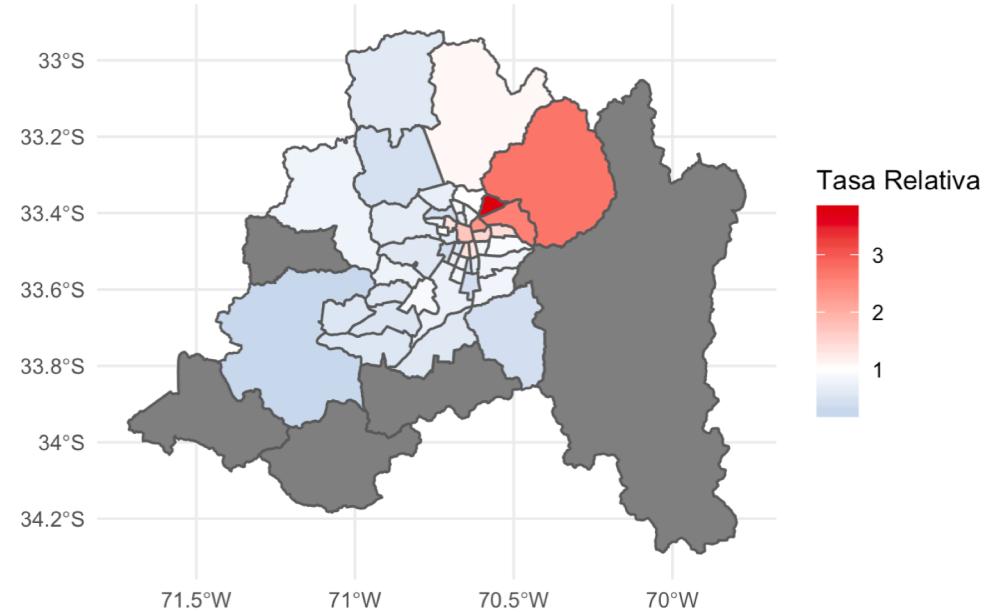
Tasa Cruda de Riesgo
Región Metropolitana - 2020-04-08



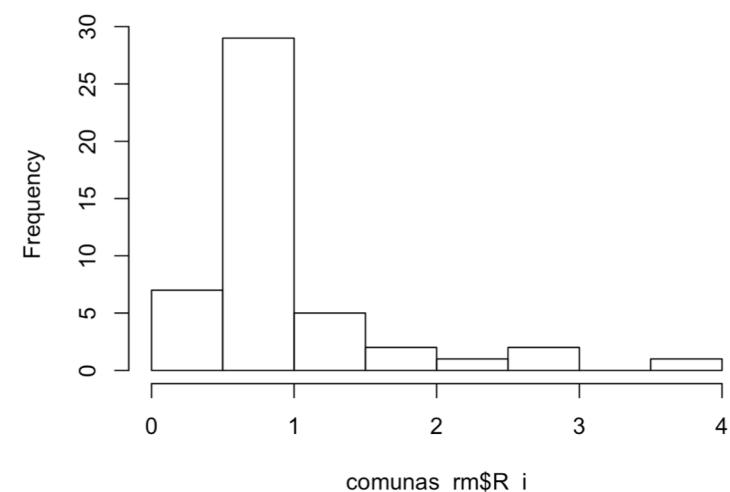
Tasa Cruda de Riesgo



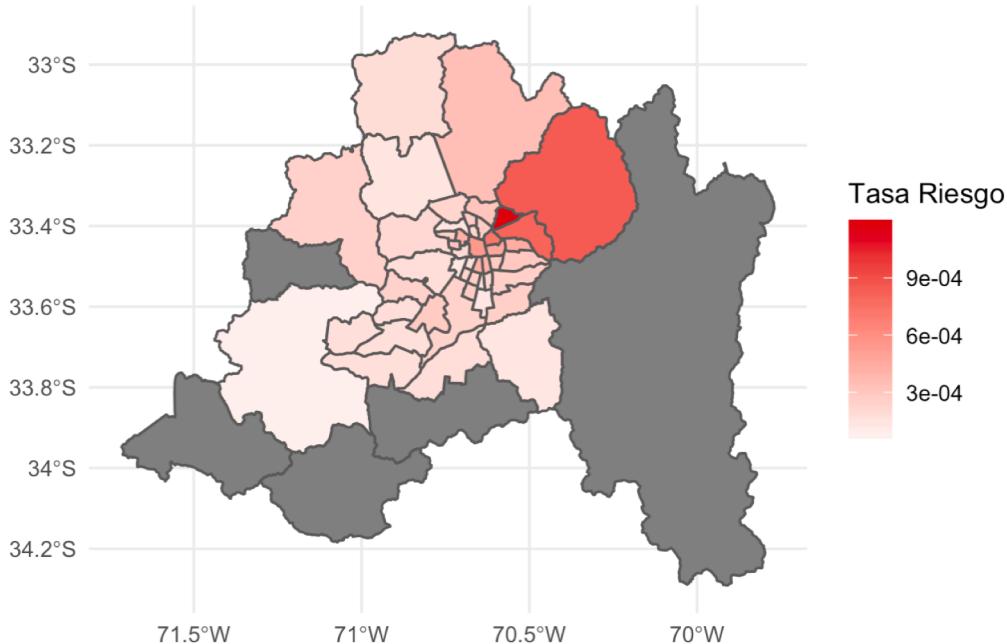
Tasa Relativa de Riesgo
Región Metropolitana - 2020-04-08



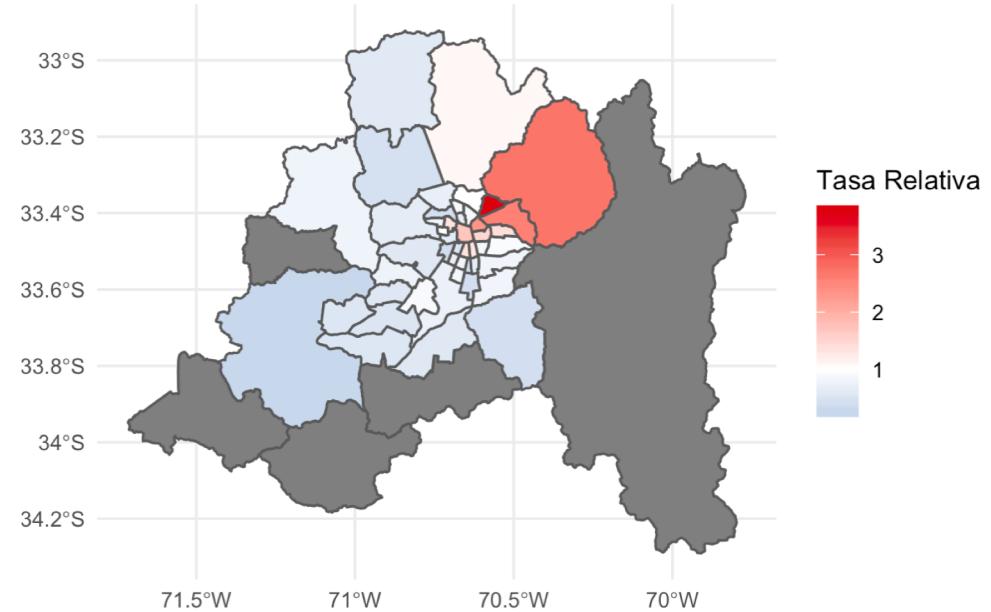
Tasa Relativa de Riesgo



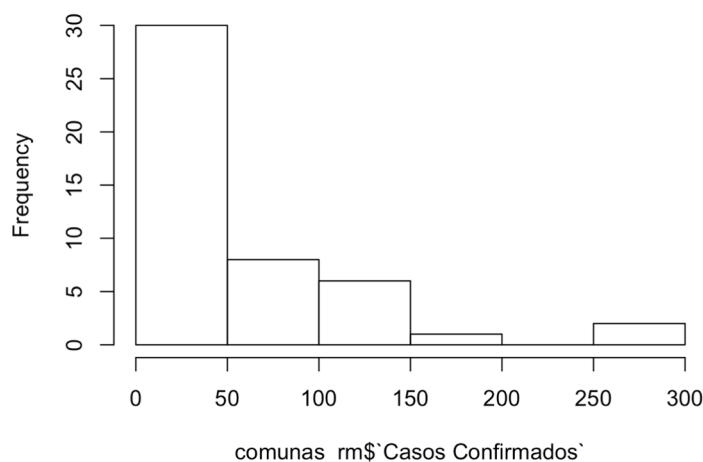
Tasa Cruda de Riesgo
Región Metropolitana - 2020-04-08



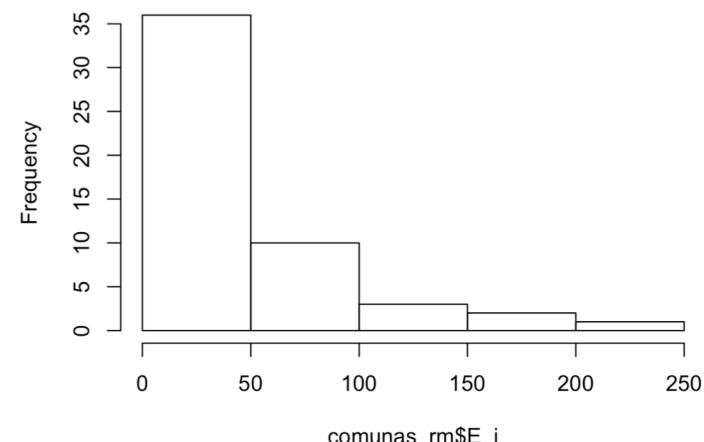
Tasa Relativa de Riesgo
Región Metropolitana - 2020-04-08



Casos Confirmados (Reportados)



Casos Esperados





Through back in time

CLASS 03

Motivation

- Substantive
 - From Adam Smith invisible hand to social networks
 - Individual vs. socio-spatial interaction
 - Peer effects, contagion, imitation, trends
 - Spatial externalities
 - Costs/Benefits from activities that impact in a different location
 - Spatial spillovers:
 - Good: neighbor playing the piano
 - Bad: neighbor with COVID-19 and not quarantining
 - Spatial multipliers
 - Investment in a park → housing prices
 - Spatial Mismatch/Disparities
 - Spatial Context

Motivation

- Practical
 - Data: geo-located observations
 - Private sources
 - Public sources
 - Self reported vs. Web scraped
 - Spatial mismatch between data and social processes
 - Labor markets vs. Cities and Counties → Labor Market areas
 - Epidemic spread vs. Data collection in hospitals
 - Neighborhood Effects
 - Spatial interpolation
 - Change of support problem
 - Data at different spatial levels that don't overlap
 - Ej: Parents' income aggregated at the school level used to predict municipalities' income level

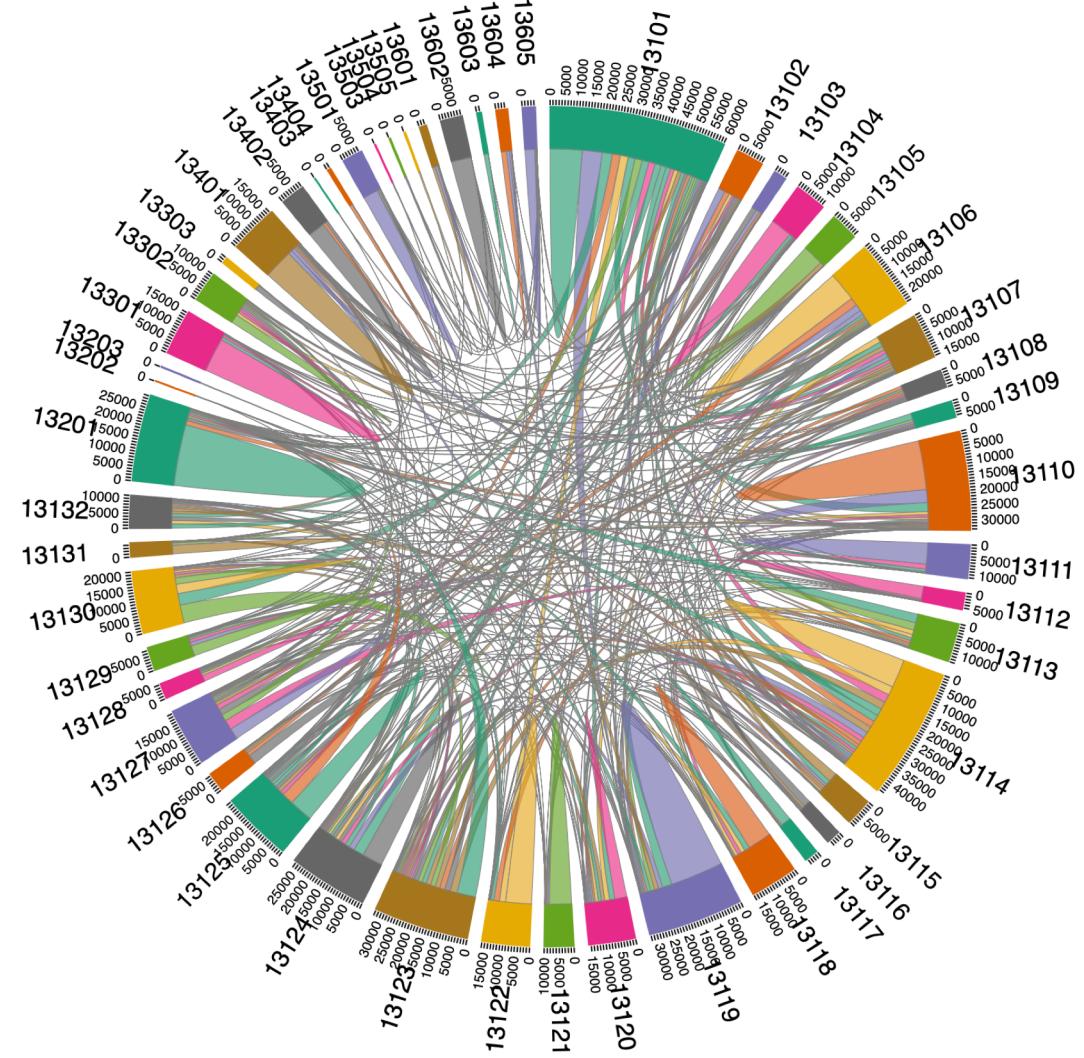
- Where do things happen
 - Patterns, clusters, hot spots, disparities,..
- Why do things happen
 - Location decisions
- How, things that happen, affect other things (spillovers) and how context affect what happens (interaction)
- Where should things be happening/be located
 - Optimization

What if we start,..

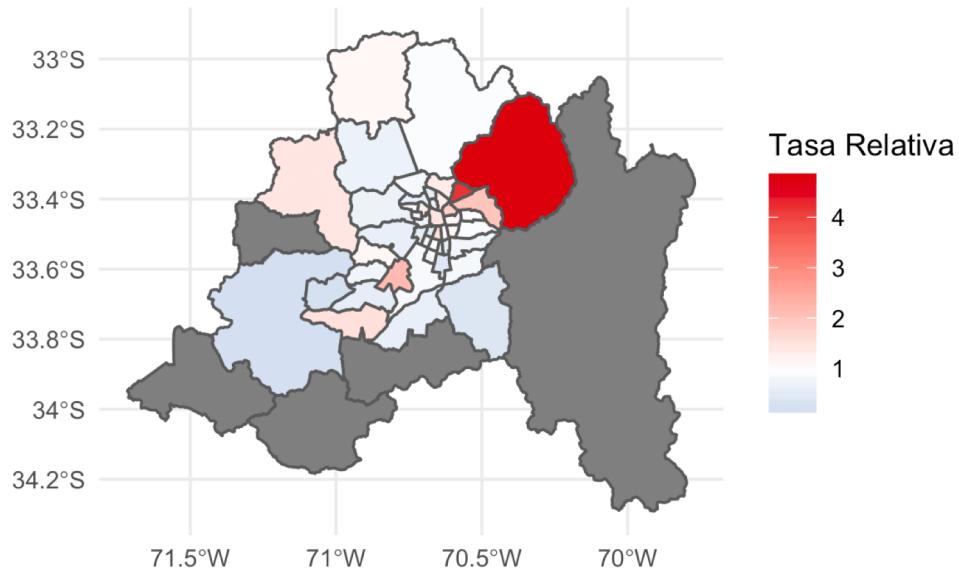
CONSIDERING WORKER'S MOBILITY

Commuting Patterns

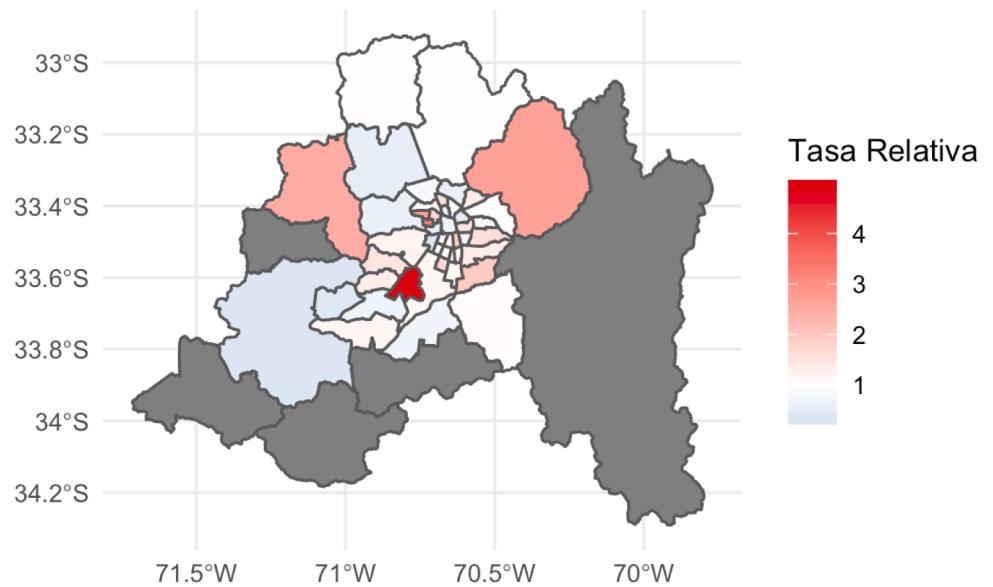
- **Commuting**
 - Traveling Process to go from home to work
 - Variables involved: distance, time, route, travel mode, decisions, etc.
 - Workers live in one municipality but work in a different one.



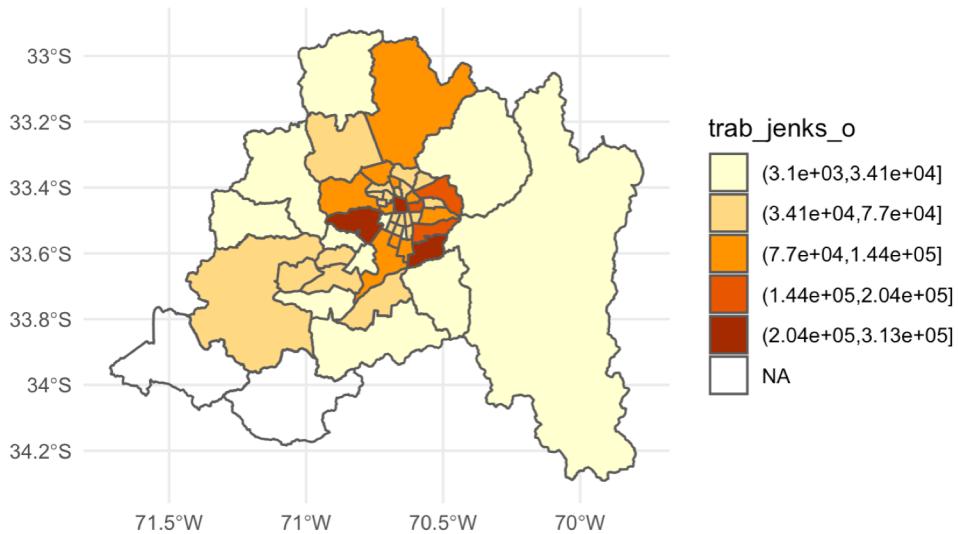
Tasa Relativa de Riesgo - Trabajadores Origen
Región Metropolitana - 2020-04-08



Tasa Relativa de Riesgo - Trabajadores Destino
Región Metropolitana - 2020-04-08



Trabajadores por Origen
Región Metropolitana - 2020-04-08



Trabajadores por Destino
Región Metropolitana - 2020-04-08

