# CaRtography: Creating Accurate and Beautiful Maps in R

By: Ayanthi Gunawardana

New York R Conference

July 13, 2023




https://images-na.ssl-images-amazon.com/images/I/51iF%2BpUk3wL._AC_.jpg

# Overview

- About Me
- Geospatial Data Overview
  - Terminology
  - Data Types
  - Projections and Coordinate Systems
- Maps Overview
  - Visual Map Elements
  - Bad Map Samples
- Maps in R
  - Packages
  - Examples of maps in R
- Resources

# Who Am I?



- B.A. in Sociology, Masters in City Planning
- Spent 8 years working at the intersection (pun intended) of transportation and data at two NYC city agencies
- Taking a break from city planning as a career, but am still interested in geospatial data and maps
- R-Ladies NYC Board Member
  - https://www.meetup.com/rladies-newyork/




https://www.researchgate.net/profile/Rasika_Muthucumarana/publication/261016208/figure/fig1/AS:717197353689089@1548004582193/Map-of-India-and-Sri-Lanka-with-the-locations-of-major-archaeological-sites.png

# Geography: One of the Oldest Sciences

- A map is probably the first data visualization you interacted with; maps are in every classroom
- The geospatial field is one of the oldest forms of science
  - Finding the curve of the Earth
  - Maps in the Age of Exploration
- Some current geospatial data applications
  - Siting for land uses (farms, retail, etc)
  - Space exploration
  - Military operations
  - Population patterns



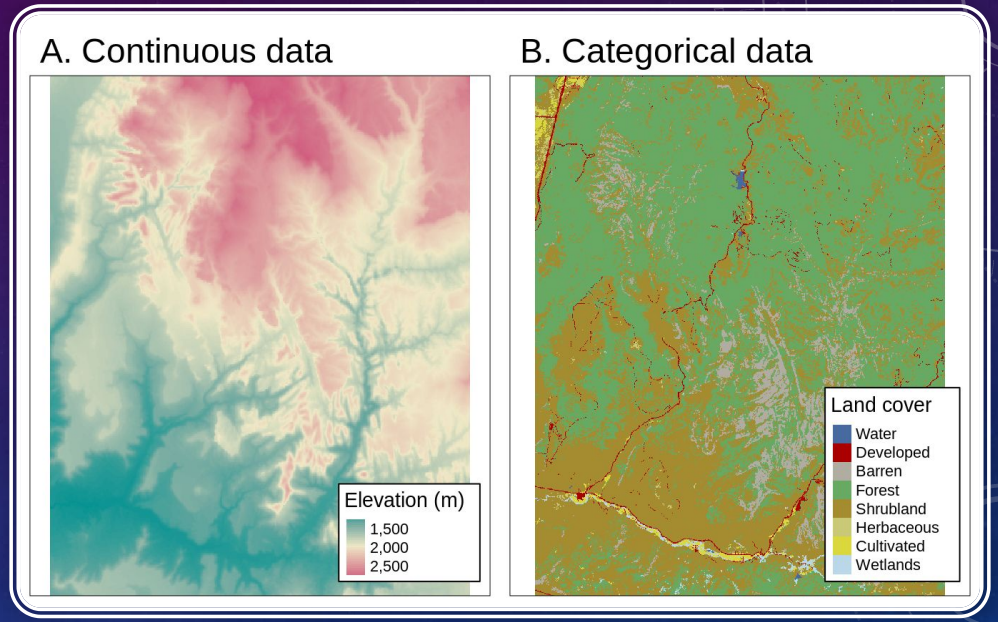

https://upload.wikimedia.org/wikipedia/commons/1/15/Old-world-map.jpg

# Mapping Terminology

- Geographic Information System (GIS):
  - a framework for gathering, managing, and analyzing data. It analyzes spatial location and organizes layers of information into visualizations using maps and 3D scenes. GIS reveals deeper insights into data, such as patterns, relationships, and situations—helping users make smarter decisions.
  - (https://www.esri.com/en-us/what-is-gis/overview)
- Cartography:
  - the science or practice of drawing maps
- Shapefile:
  - Vector data format for GIS software.
  - Represents points, polygons, or lines; contains attribute and geospatial information about each feature
  - As GIS becomes more open source, other formats are becoming more popular, like Geopackage, GeoJSON, KML
  - GIS data can also be stored in csv files, but requires more manipulation in R

# Vector Data vs. Raster Data

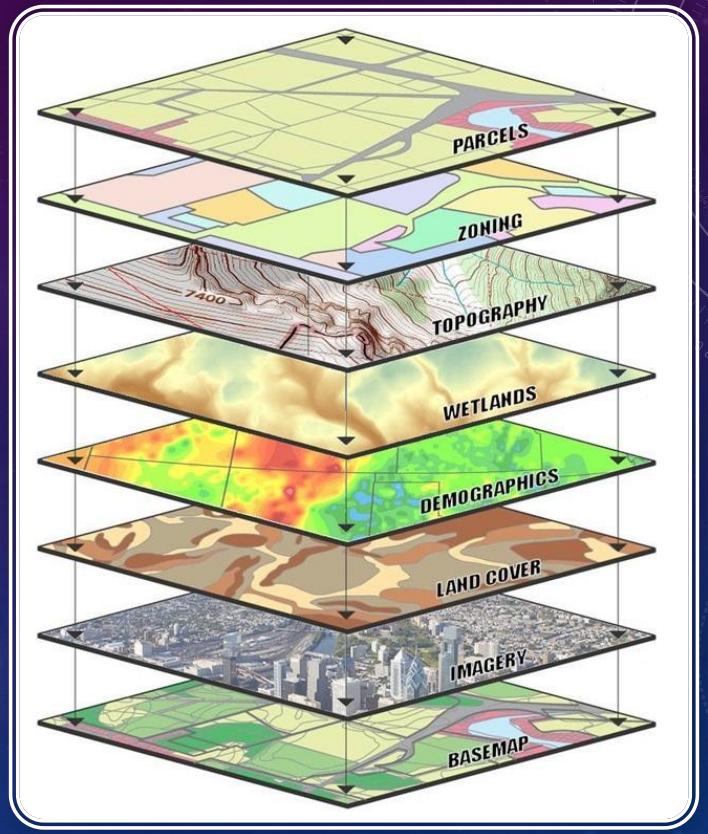
- Vector Data
  - Consists of points, lines, and polygons
  - Each feature type needs to be stored separately
- Raster Data
  - Array of cells (pixels) used to represent real-world objects; a digital image
  - Each pixel has its own value
  - Example: orthography
    - aerial images of land that are geometrically corrected so the scale is uniform
  - Example: digital elevation model (DEM)
    - 3D representation of the Earth's surface using Lidar, radar, and other remote sensing tools




https://geocompr.robinlovelace.net/spatial-class.html#raster-data

# Data Representation

- Different types of data are represented in different layers
- Using these layers in a series of maps allow us to see the “whole picture”
- Spatial information
  - Geometry
  - Location
- Non-spatial information
  - Attribute information for each feature
  - Example: population in a specific census tract; population itself is not spatial, but is aggregated to a spatial entity




https://www.usgs.gov/media/images/gis-data-layers-visualization

# Geographic Coordinate System vs. Projected Coordinate System

- "Maps are scale models of reality" – Mark Monmonier
- A map is a visual representation of reality, and is **inherently distorted**
  - You're taking a piece of the globe and flattening it to create a map
  - Think about peeling an orange and how different the peel looks when flattened
- A geographic coordinate reference system describes where a location is on Earth
  - Since the Earth is round, this is measured in angular units, usually degrees
- A projected coordinate system is a geographic coordinate reference system that has been flattened
  - When using a projected coordinate system, you can measure distances using a straight line

# Conic Projection

- The point where the cone touches the globe is known as the standard parallel
- Since a cone is wider than it is tall, the map becomes more distorted as you move north or south of the standard parallel

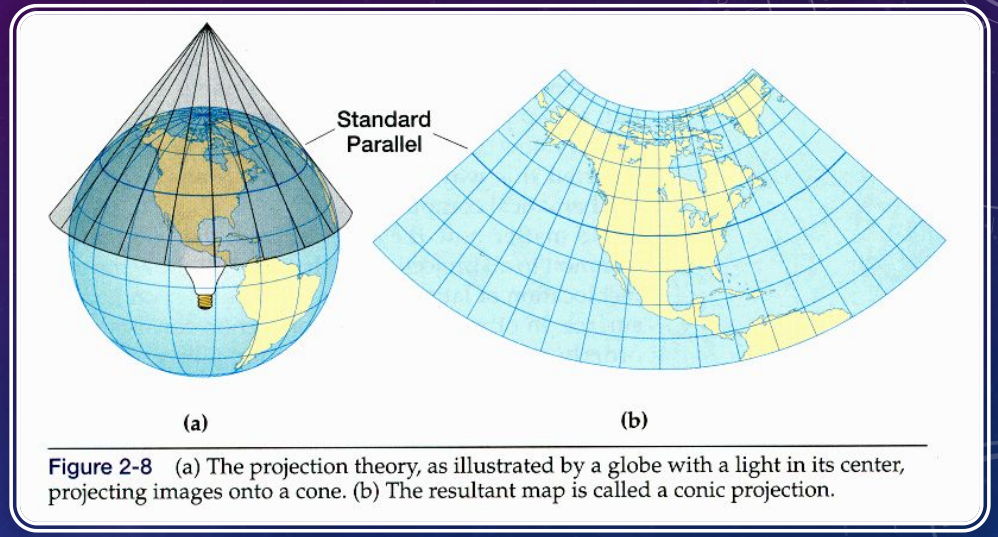
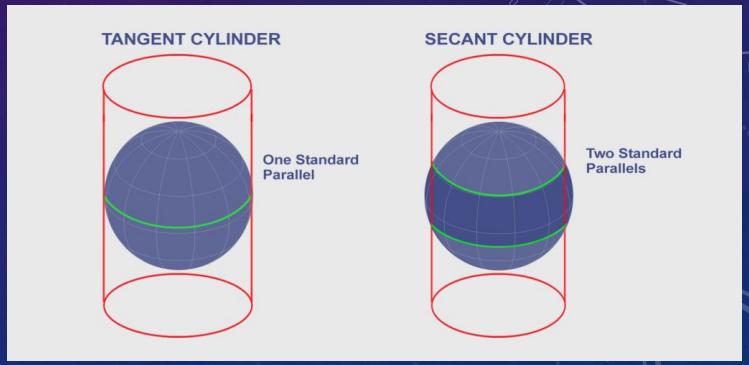



Figure 2-8 (a) The projection theory, as illustrated by a globe with a light in its center, projecting images onto a cone. (b) The resultant map is called a conic projection.

http://faculty.kutztown.edu/courtney/blackboard/Physical/05Project/conproj.html

# Cylindrical Projection

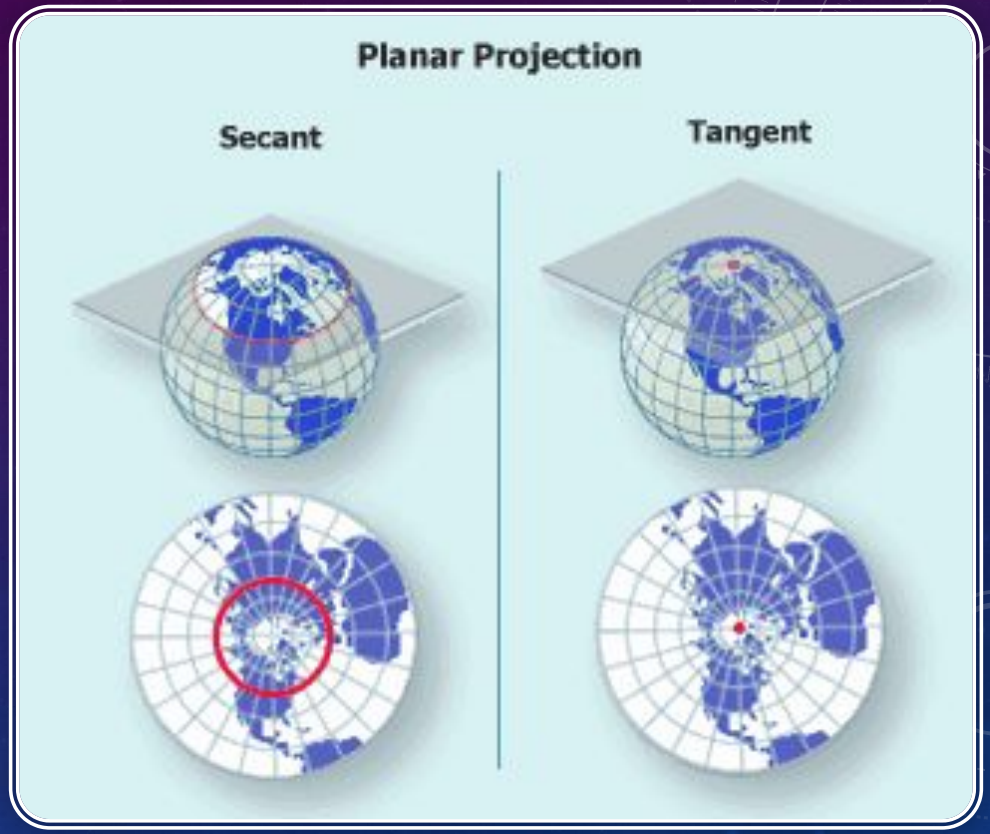
- Commonly used for world maps
- The globe is represented as a cylinder that is tangent or secant to a selected line, and then flattened
- Tangent vs. secant (http://geokov.com/education/map-projection.aspx)
  - Tangent: cylinder's circumference touches the reference globe's surface;.
  - Secant: cylinder intersects the globe; At the place where the cylinder cuts through the globe two secant lines are formed.
- The scale is constant at the tangent and secant lines, but the farther away you get from these lines, the more distorted the map becomes




https://gisgeography.com/cylindrical-projection/

# Planar Projection

- Commonly used to highlight a specific area
- This plane is tangent to the globe at a single point
- Good for preserving distances
- However, distortions increase as you increase the size of the map and move away from the center



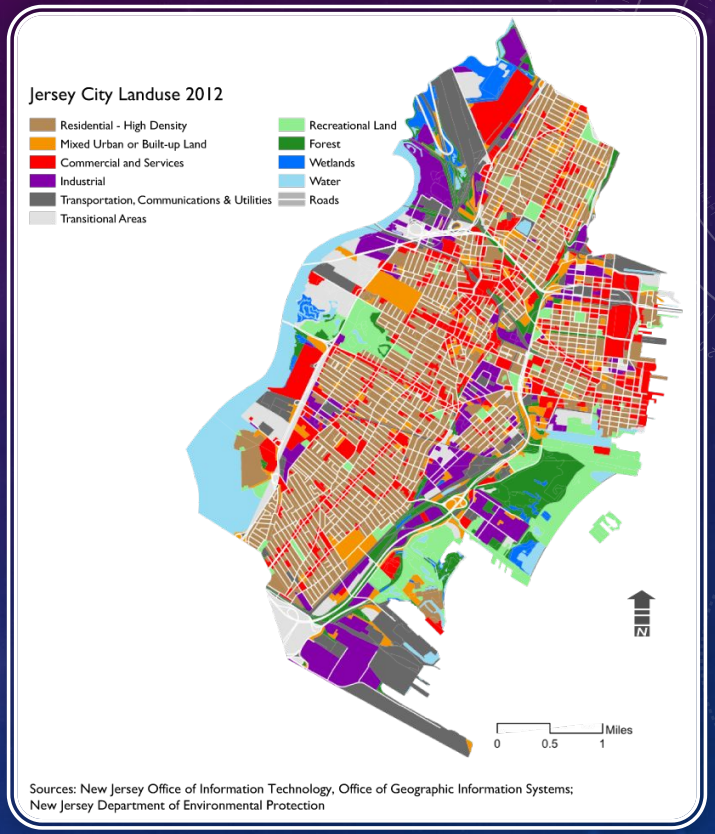

USGS

# Best Practices in Visualizing Geospatial Data

- Who is your audience?
- Always normalize your data
  - Unequal geographic boundaries, areas, and populations can make data seem concentrated when it's just a reflection of the population density
- Keep labels to a minimum
  - Over-labeling makes a map hard to read
- Think critically about your color palette
  - Does your organization have its own palette?
  - Accounting for color-blindness
  - Industry standards
- Document all changes to the source data

# Visual Map Elements

- Essential Elements
  - Map
  - Legend
  - Title
  - Data sources
- Optional (depends on the map)
  - North arrow
  - Scale bar
  - Labels




http://njbikeped.org/jersey-citys-new-citi-bike-stations-how-to-benefit-all-neighborhoods-equally/

# What Do We Do With All This Data?

We can use any combination of this data to make a good map, but using too much data or the wrong type of data will ruin a map!

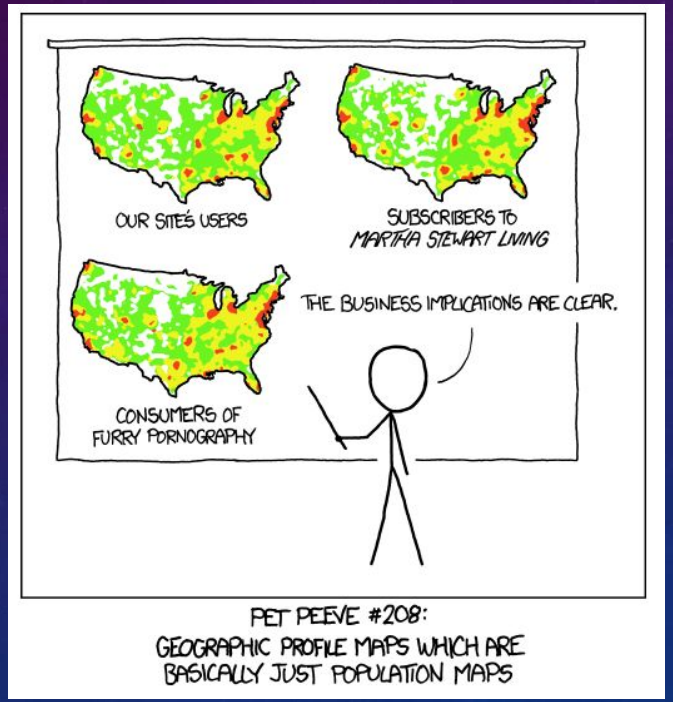
The following slides contain examples of bad and/or misleading maps and cartography practices.

# Fun, But Chaotic



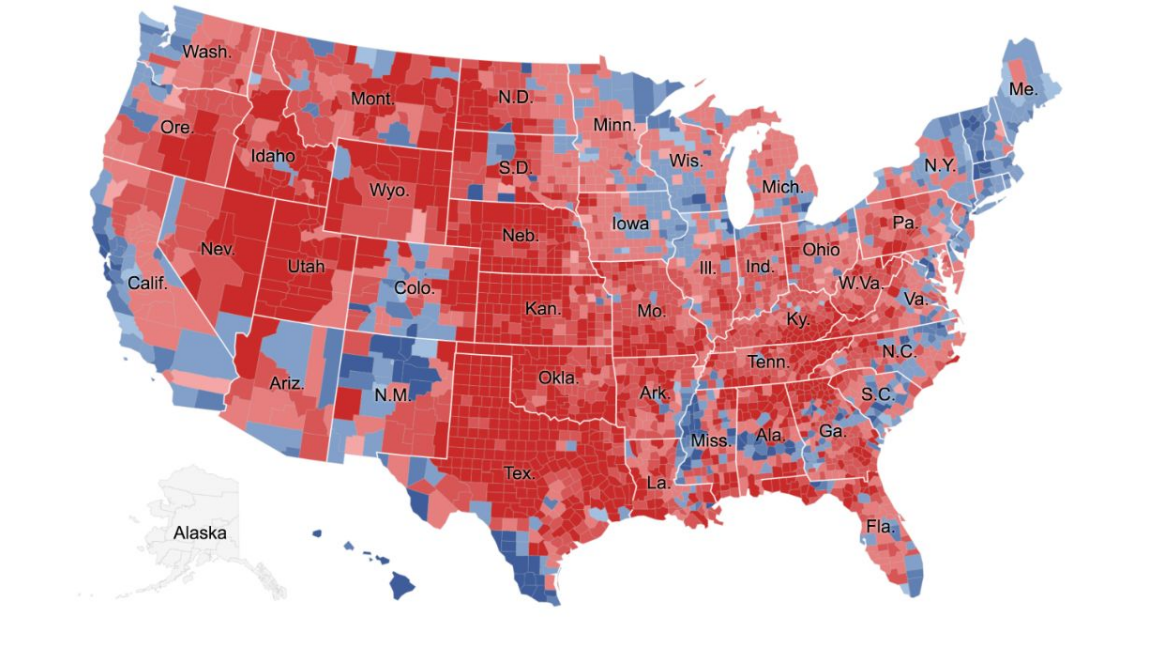

https://www.reddit.com/r/terriblemaps/comments/fj7ih8/us_state_map_marked_with_a_tv_show_that_takes/

# Sometimes It's Just a Population Map



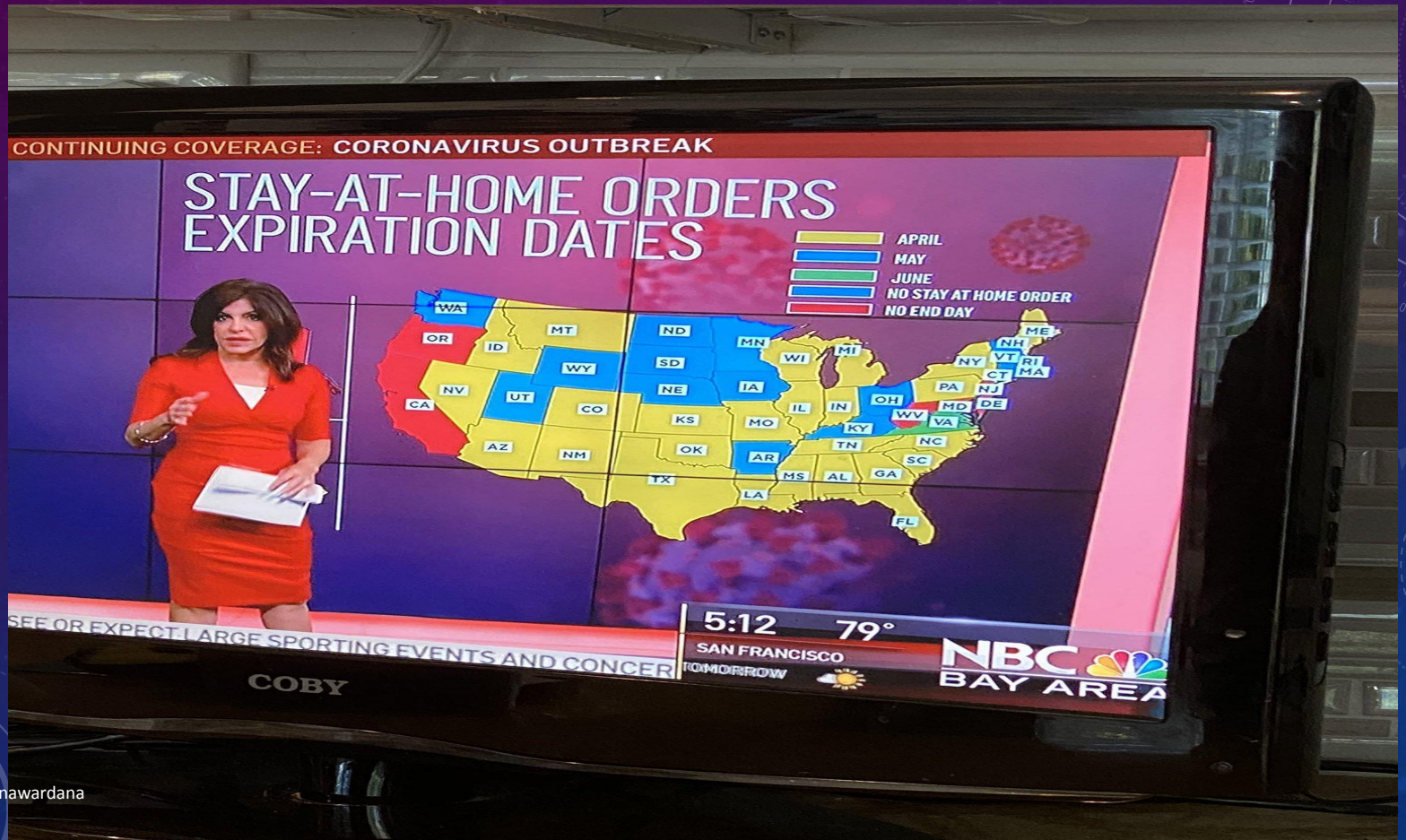

xkcd

# Misleading Election Maps

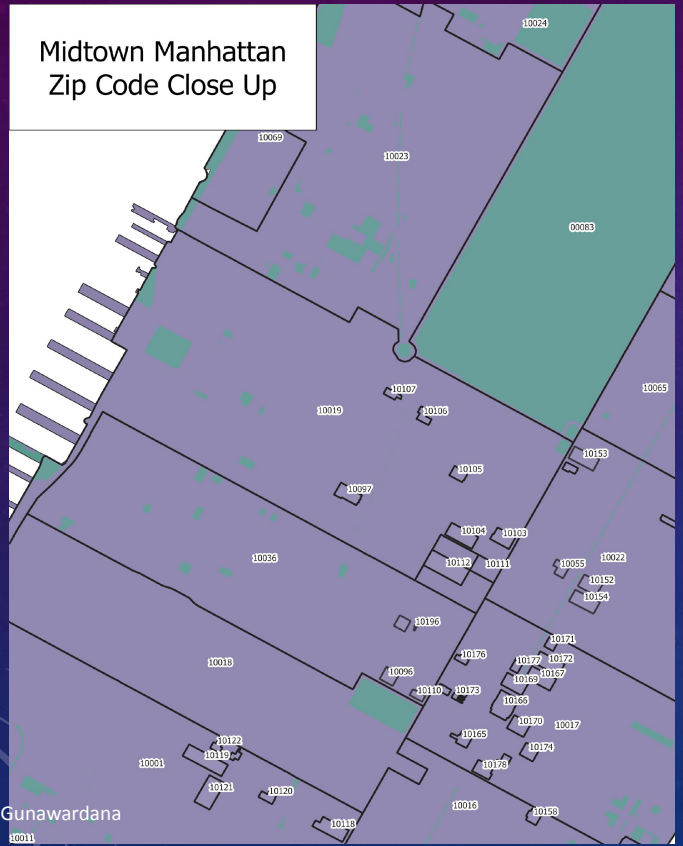


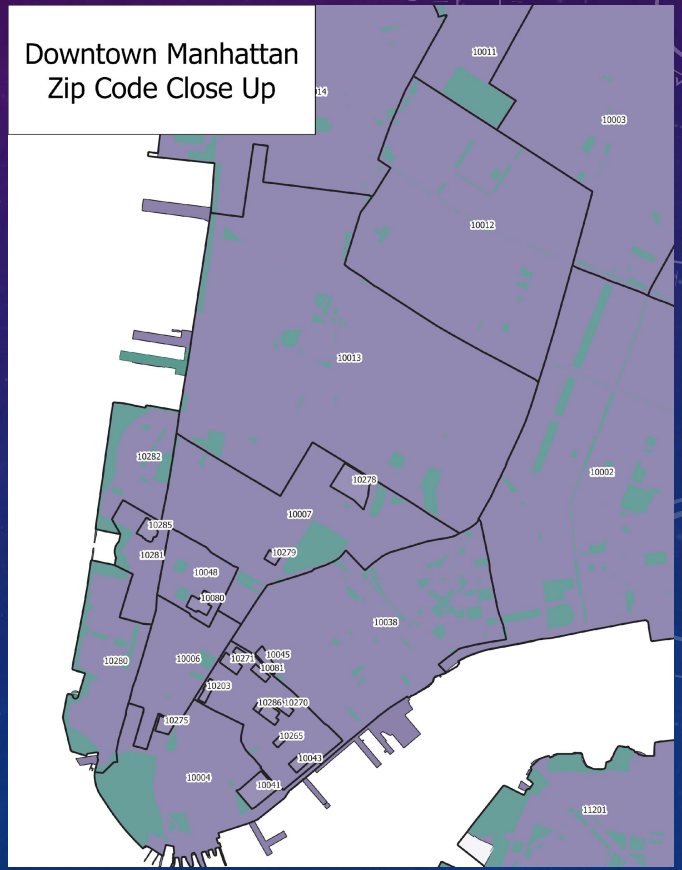

https://www.nytimes.com/interactive/2016/11/01/upshot/many-ways-to-map-election-results.html

# Just....BAD


CONTINUING COVERAGE: CORONAVIRUS OUTBREAK
STAY-AT-HOME ORDERS EXPIRATION DATES
APRIL
MAY
JUNE
NO STAY AT HOME ORDER
NO END DAY
WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
MD
DE
NJ
CT
RI
MA
NH
VT
ME
SEE OR EXPECT LARGE SPORTING EVENTS AND CONCER
5:12
79°
SAN FRANCISCO
TOMORROW
NBC BAY AREA
COBY

# The Chaos of Zip Codes



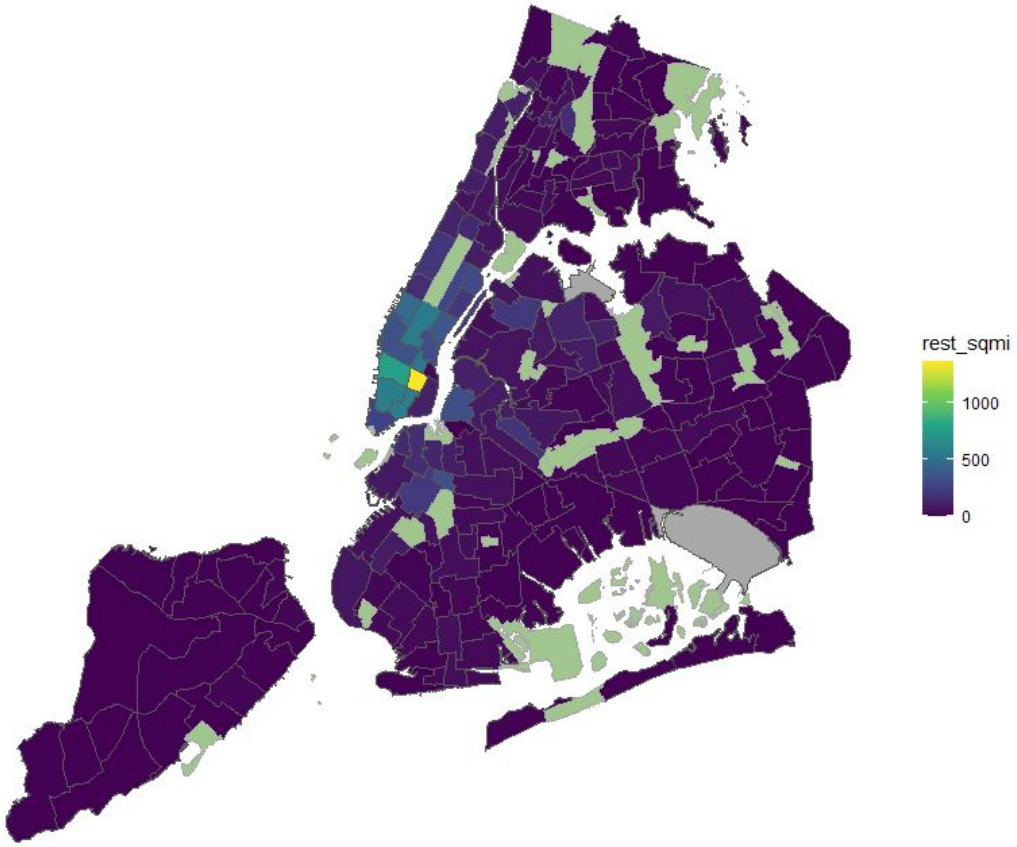
# R Packages Used to Make Maps

# Packages for Manipulating Spatial Data

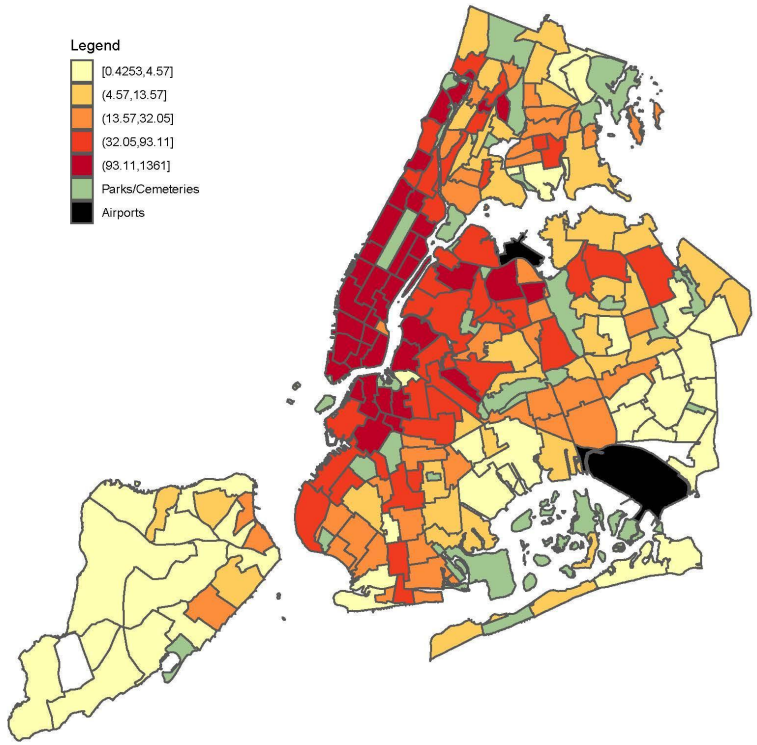
- sf: the main package for working with geometries and various spatial functions
- Ggplot2: static maps
- Mapview: interactive map viewing in the IDE
- Ggspatial: custom map tiles and labels
- Rgdal: formatting spatial file types
- Tmap: static and dynamic maps
- Cartography: create maps that look like more traditional GIS-software maps
- Leaflet: shiny apps and dynamic visuatizations
- Tidycensus: extracting and manipulating census data
- Rayshader: 3D maps
- Raster: manipulating and visualizating raster data
- Ggmap: uses google maps functionality for mapmaking
- Tigris: additional census data

# Examples of Maps Made in R

# Open Restaurant Permits Per Square Mile
# New York City




## Open Restaurant Permits
## Per Square Mile
## by Neighborhood Tabulation Area




source: NYC Open Data, Retrieved on March 8, 2022

Source:
https://milospopovic.net/maps/








https://twitter.com/researchremora/status/1567930733087399936


NEW JERSEY
Atlantic City

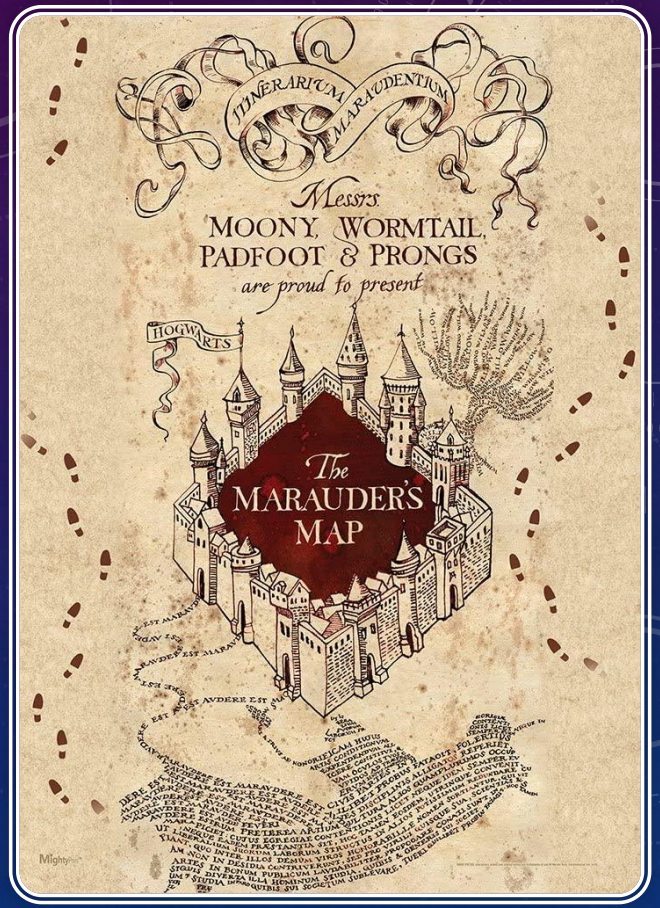

PATERSON
East Orange
NEW YORK
Elizabeth
Lower New York

# Useful Resources

- Books
  - “How to Lie With Maps” by Mark Monmonnier
  - “Geocomputation in R” by Robin Lovelace, Jakub Nowosad, and Jannes Muenchow https://r.geocompx.org/
- Geospatial Professionals with Educational Content
  - Matt Forrest: CARTO employee https://forrest.nyc/
  - Rafael H.M. Pereira: creator of some spatial R packages https://www.urbandemographics.org/about/
  - Milos Popovic: aesthetically pleasing maps in R https://milospopovic.net/
  - Kate Berg: environmental analysis in R: https://www.pokateomaps.com/
  - Robin Lovelace: big contributor to R geospatial community: https://www.robinlovelace.net/
- NYC Open Data Portal – has several geospatial datasets to get you started with analysis
  - https://opendata.cityofnewyork.us/

# Thank You for Listening – Mischief Managed!

Github: https://github.com/argunaw




https://images-na.ssl-images-amazon.com/images/I/81hXyIewnxL._AC_SL1000_.jpg