Assessment & Evaluation of Surveillance Programs

Dr. Rachel Sippy
University of St Andrews

Objectives

- Review attributes of a surveillance system
- Discuss assessment for surveillance systems

Post Questions in the Chat!

(we will have breaks to answer these during the workshop)

Workshop Schedule

Time	Topics
2:00—2:15 pm	Surveillance Attributes
2:15—2:45 pm	System Assessment
2:45—3:00 pm	Example Assessment
3:00—3:10 pm	Break
3:10—4:00 pm	Mapping

Surveillance Systems

Attributes

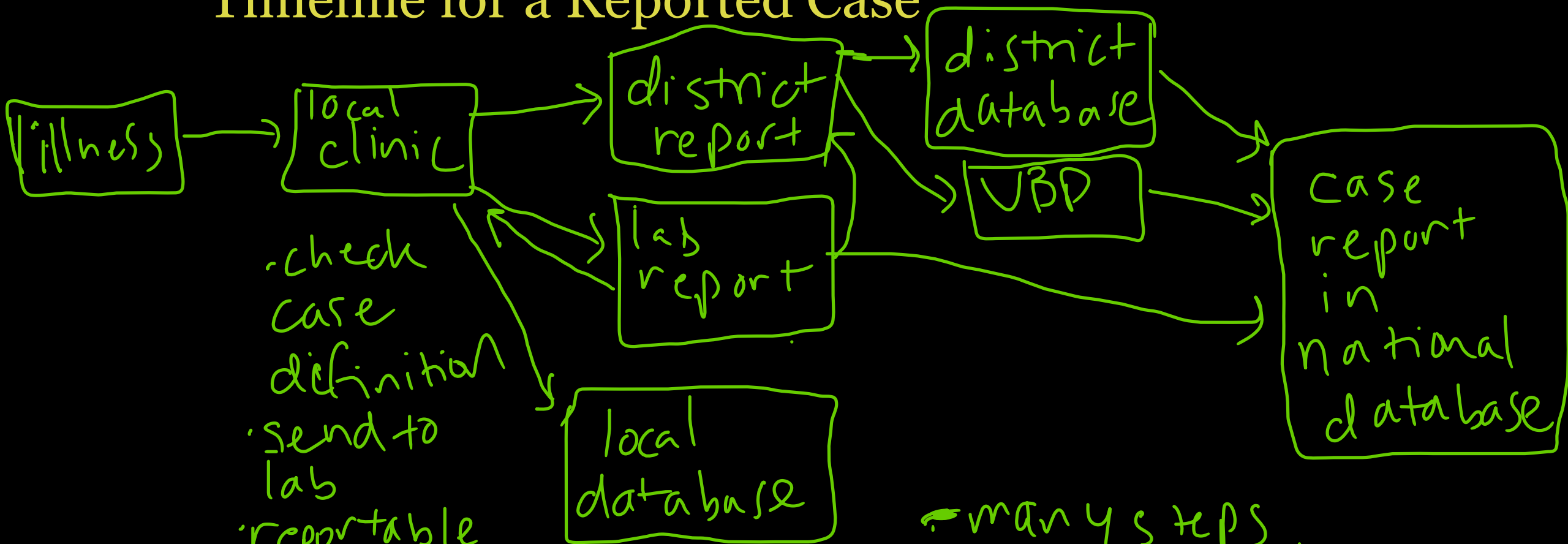
- Useful
- Reliable, high-quality data
- Timely
- Flexible
- Simple
- Stable
- Sensitive
- Positive predictive value
- Representative
- Acceptable

Attributes

- Continuous, ongoing data collection
- Efficient, practical, timely
- Flexible, acceptable
- Sensitive, representative
- Direct link between outputs and action

Zika

Timeline for a Reported Case



- check case definition
- send to lab
- reportable illness!!

- many steps
- potential delays

System Assessment

Purpose of Evaluations

- assessment of system performance and its attributes
- with assessment, able to:
  - ensure the systems meets its objectives effectively as possible
  - increase system utility and efficiency
  - provide recommendations for the future

Assessments

- should be conducted regularly
- may occur at country, regional, international level
- WHO and technical partners are involved in regional and international assessments
  - specific indicators used to assess surveillance systems

Indicators

- different from health indicators
- should be measurable and action-oriented
- country or WHO may set targets for these indicators
- specific to a disease or set of diseases
- related to the ideal attributes of a surveillance system
  - completeness: proportion of facilities and districts reporting the disease
  - timeliness: proportion of reports sent on time

- target for well-performing surveillance 80%

- immediate
  w/in 24 hours
- week/month

%

facilities → clinics

%

→ district

Example: Polio Surveillance

Completeness, representativeness of reporting	• At least 80% of expected AFP reports on time, no reporting where cases were not followed up. Representative sites.
Sensitivity of surveillance	• At least one case of non-polio AFP detected annually per 100000 people <15
Completeness of case investigation	• All AFP cases with full clinical and virological investigation with at least 80% having adequate stool specimens
Completeness of follow-up	• At least 80% of AFP cases should have follow-up examination for residual paralysis at 60 post onset
Laboratory performance	• All AFP case specimens must be processed in a WHO-accredited laboratory within the Global Polio Laboratory Network

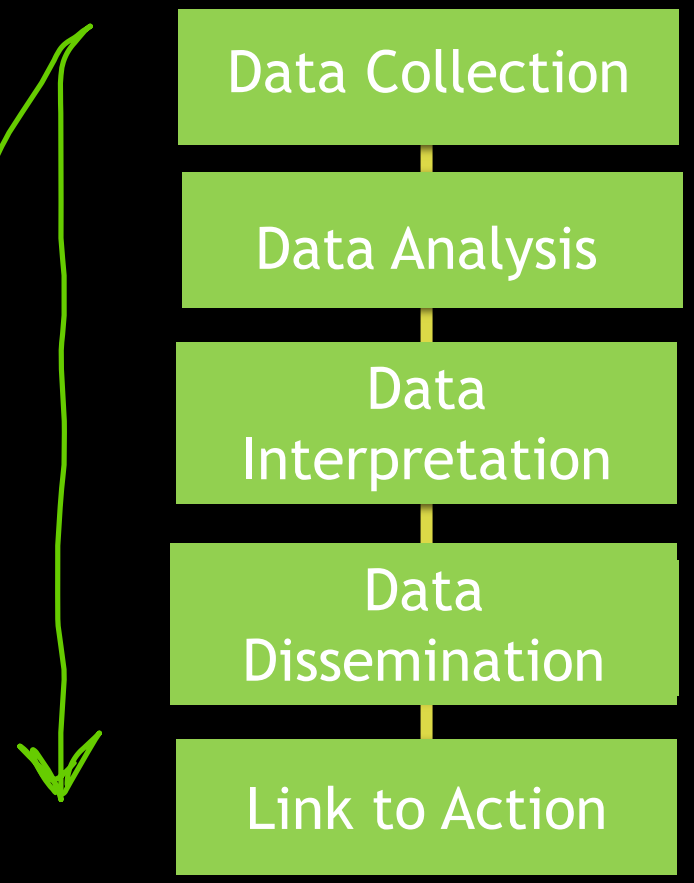
Example: Indicators for Comprehensive VBD Surveillance

data quality completeness	• % of districts reporting any VBD in the last 12 months
lab capacity	• % of countries with access to laboratory capacity to test for at least one VBD
? → are you detecting rashes?	• % of countries achieving non-measles/non-rubella discard rate of ≥2/100000 people in a 12-month period
? → are you detecting AFP?	• % of countries achieving the non-polio acute flaccid paralysis rate of >1/100000 among <15 years population in a 12-month period

Comprehensive Evaluation of Surveillance Systems

Approaches to Evaluation

- Entire surveillance system
- Surveillance of individual diseases or sets of diseases
- Components of the surveillance process



System Evaluation

1. Describe the system in detail
2. Use this information to create a directory of activities and how they relate to policy and actions
3. Identify gaps
4. Identify priorities for development

System Evaluation

1. Describe the system in detail
2. Use this information to create a directory of activities and how they relate to policy and actions
3. Identify gaps
4. Identify priorities for development

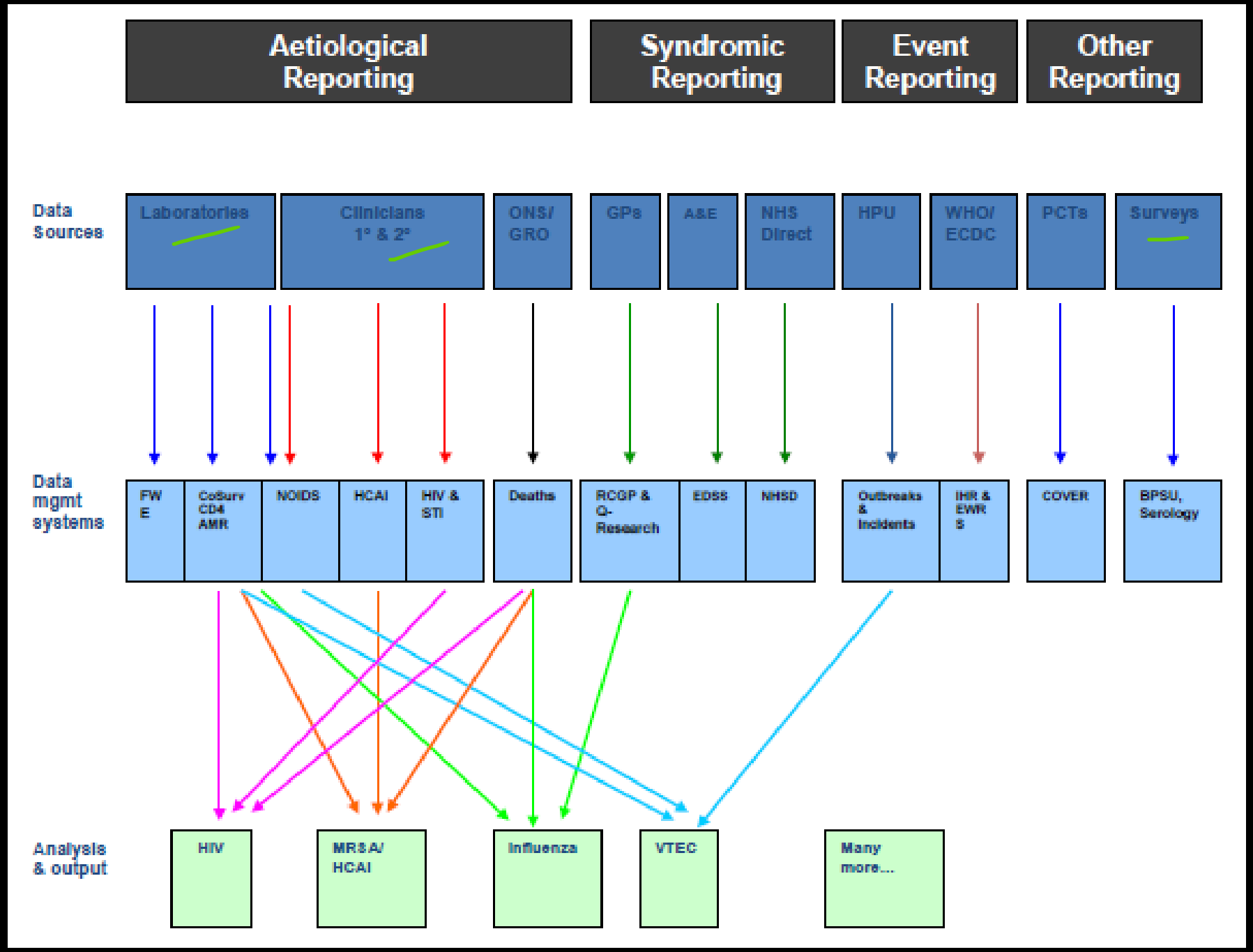
- What is the system structure?
- Who is responsible for outputs?
- Who/where are the reporting sites?
- What are data reporting guidelines?
- What is being monitored?
- What are the minimum data collected?
- What is the deadline for reporting?
- What are the objectives?

System Evaluation

1. Describe the system in detail
2. Use this information to create a directory of activities and how they relate to policy and actions
3. Identify gaps
4. Identify priorities for development

- What databases store information?
- How do data get to databases?
- Who performs analysis?
- Who performs interpretation?
- How are results disseminated?
- Who makes action decisions?
- What is the deadline for reporting?
- What are potential actions?

data flow England



← types of surveillance

← generate data

← data bases

← analysis with data

Data item description	Source type	Examples of where information would be used (not exhaustive)
Cancer incidence/ mortality/ survival	Cancer Registry	Monitoring 3 yearly rolling average incidence rates for different cancer sites, at PCT and local area level, etc.
Cancer screening data	Cancer Screening	Screening analysis - uptake and coverage.
Congenital Abnormalities	Congenital Anomaly Register	
Seasonal influenza vaccination uptake	Government Department	
British Crime Survey	Government Department	WHIST Programme indicator
The English Housing Survey - EHS	Government Department	WHIST Programme indicator
Communicable Diseases	HPA Local	To inform JSNAs
Communicable Diseases	HPA National	To inform JSNAs
Immunisations	HPA National	To inform JSNAs to inform HPV programme
Enhanced Infectious Disease Surveillance e.g. STI/HIV, TB, HCAI	HPA National	To inform JSNAs

3

link to action

System Evaluation

1. Describe the system in detail
2. Use this information to create a directory of activities and how they relate to policy and actions
3. Identify gaps
4. Identify priorities for development

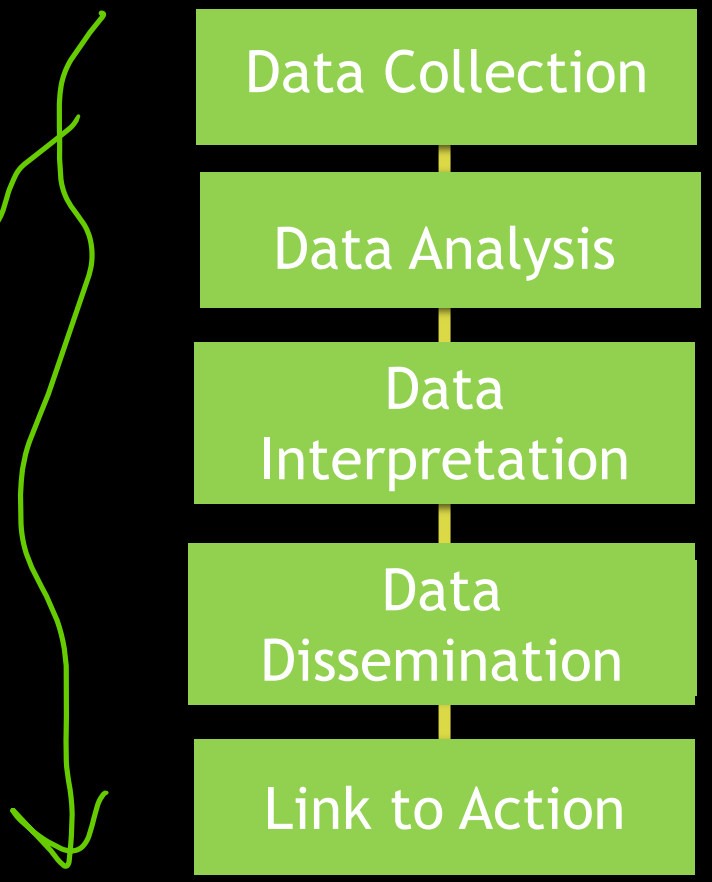
new disease?
surveillance?
data?

integrating

- additional resources
- new guidelines: common methods, analysis, interpretation

Evaluation of Process Components

- Entire surveillance system
- Surveillance of individual diseases or sets of diseases
- Components of the surveillance process



Evaluation of Process Components

- data standards (content/code)? quality control?
- regularity, sensitivity, representativeness

Data Collection

- what analysis is performed?
- timeliness, efficiency, practicality

Data Analysis

- practical, direct link to actions

Data Interpretation

- timeliness, regularity, practicality, acceptability, direct link to action

Data Dissemination

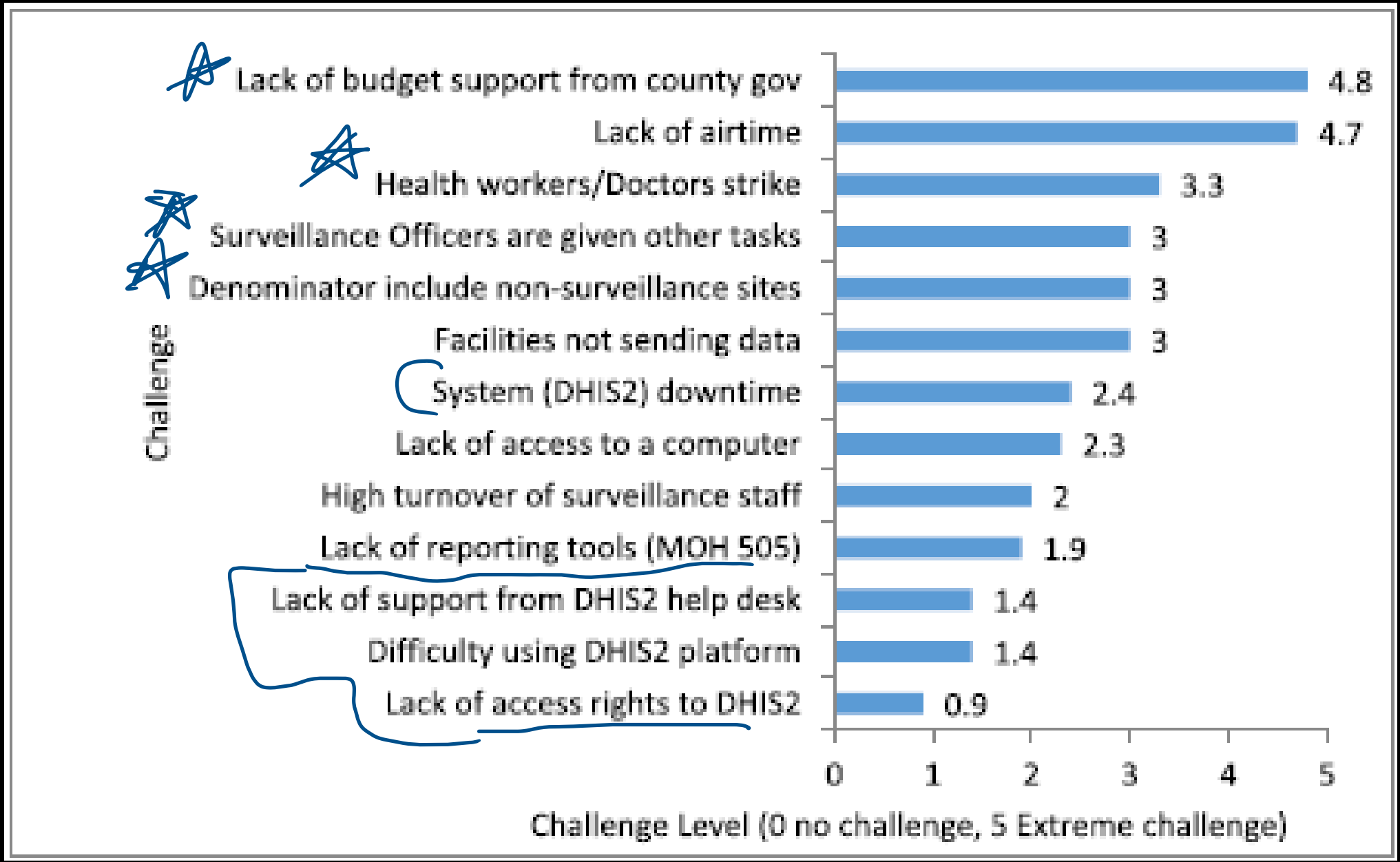
- what are the action guidelines?
- timeliness, acceptability, direct link to action

Link to Action

Evaluation of Process Components

Survey reporters

• is this going well?	Data Collection
• should this be done locally?	Data Analysis
• should this be done locally?	Data Interpretation
• is this going well?	Data Dissemination
• should this be done locally?	Link to Action


Lack of budget support from county gov
4.8
Lack of airtime
4.7
Health workers/Doctors strike
3.3
Surveillance Officers are given other tasks
3
Denominator include non-surveillance sites
3
Facilities not sending data
3
System (DHIS2) downtime
2.4
Lack of access to a computer
2.3
High turnover of surveillance staff
2
Lack of reporting tools (MOH 505)
1.9
Lack of support from DHIS2 help desk
1.4
Difficulty using DHIS2 platform
1.4
Lack of access rights to DHIS2
0.9
Challenge
0
1
2
3
4
5
Challenge Level (0 no challenge, 5 Extreme challenge)

Evaluation of Process Components

national
↓
district
↓
local

survey users
experts @ national

need local context

survey users

local knowledge for local action

• is this going well?	Data Collection
• should this be done locally?	Data Analysis
• should this be done locally?	Data Interpretation
• is this going well?	Data Dissemination
• should this be done locally?	Link to Action

- sensitivity: are we detecting cases?

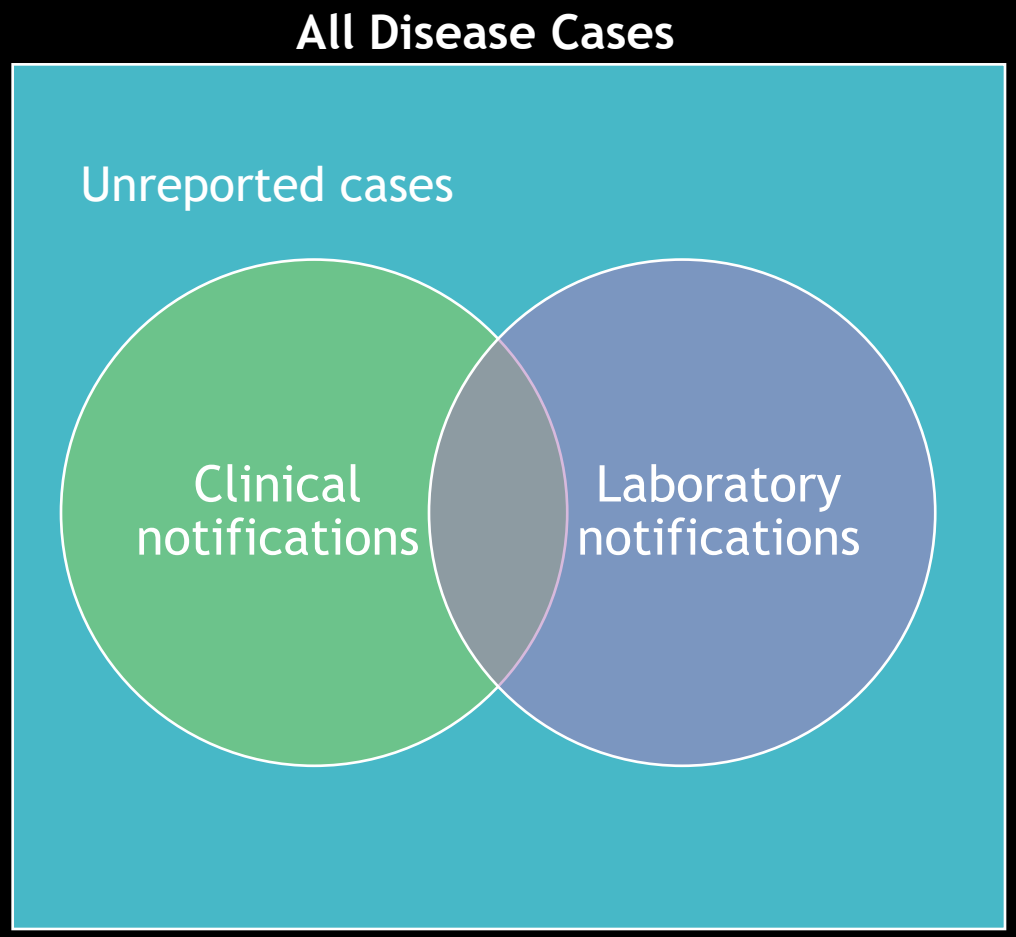
Example: Assessments for Sensitivity

Jansson *et al.* (2005) Sensitivity of the Swedish statutory surveillance system for communicable diseases 1998-2002, assessed by the capture-recapture methods. *Epidemiol Infect*, 133:401-407.

Capture-Recapture for Sensitivity

reporting sens

- goal: determine the sensitivity of the surveillance system to diagnosed cases
- method: capture-recapture
  - source A: clinical notifications
  - source B: laboratory notifications

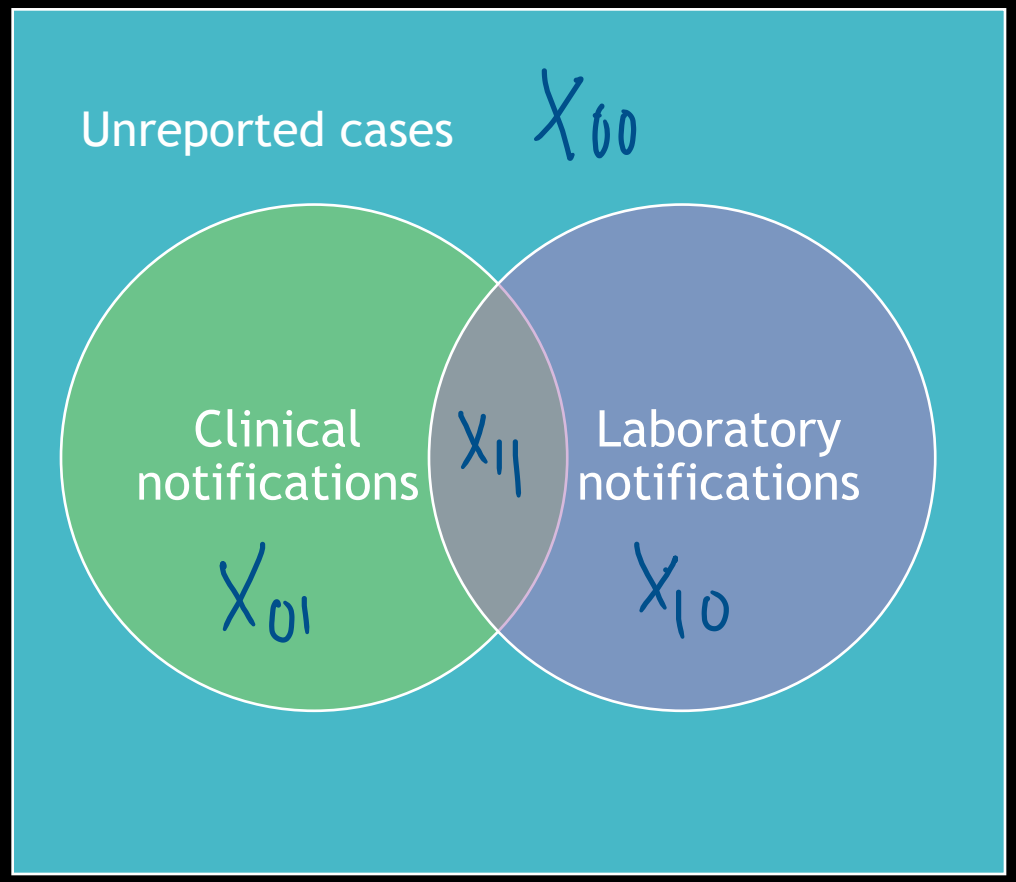


Capture-Recapture for Sensitivity

- goal: determine the sensitivity of the surveillance system to diagnosed cases
- method: capture-recapture
  - source A: clinical notifications
  - source B: laboratory notifications

$$\widehat{x_{00}} = \frac{x_{10} * x_{01}}{x_{11}}$$

?



Capture-Recapture for Sensitivity

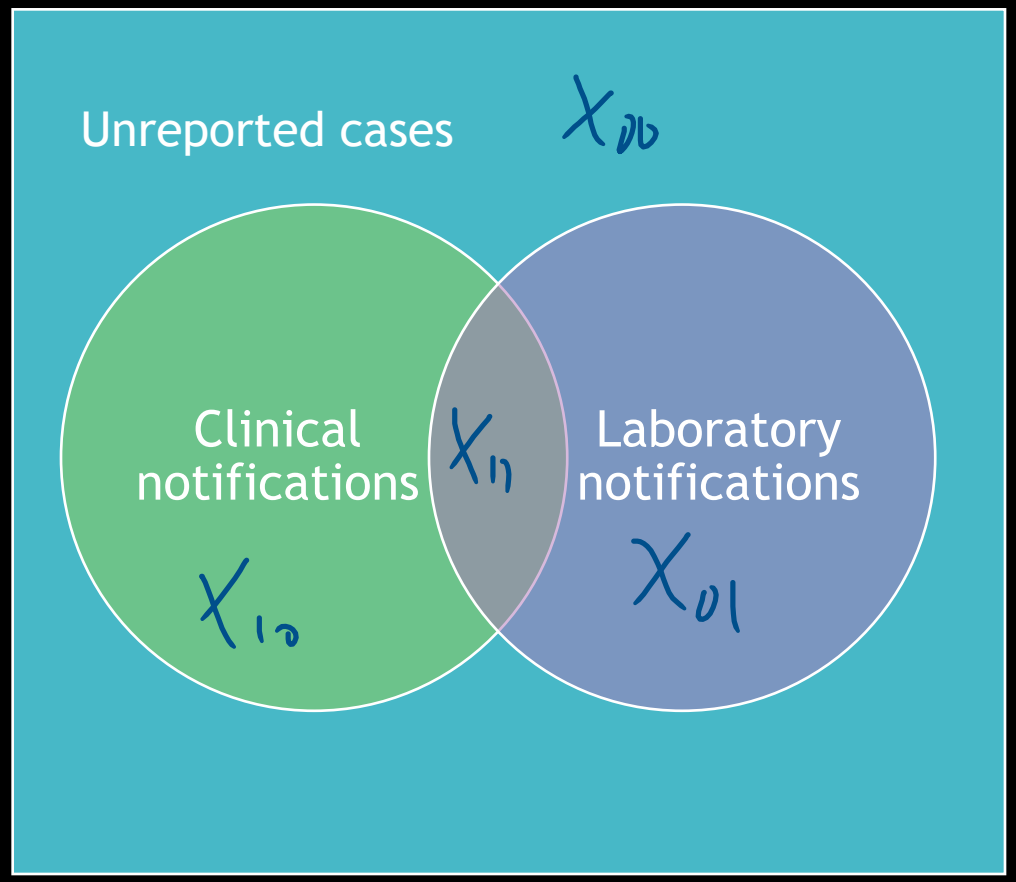
- goal: determine the sensitivity of the surveillance system to diagnosed cases
- method: capture-recapture
  - source A: clinical notifications
  - source B: laboratory notifications

$$\widehat{x_{00}} = \frac{x_{10} * x_{01}}{x_{11}}$$

- calculate sensitivity

$$\frac{x_{10} + x_{11} + x_{01}}{x_{11} + x_{00} + x_{10} + x_{01}} =$$

$$sens = \frac{reported\ cases}{total\ cases}$$



Capture-Recapture for Sensitivity

- study of surveillance system in Sweden
- data from clinical reports and laboratory reports
- four diseases reported during 1998-2002
  - invasive meningococcal infection
  - salmonellosis
  - penicillin-resistant *Streptococcus pneumoniae*
  - tularemia

Capture-Recapture for Sensitivity

- study of surveillance system in Sweden
- data from clinical reports and laboratory reports
- four diseases reported during 1998-2002
  - invasive meningococcal infection
  - salmonellosis) food borne
  - penicillin-resistant *Streptococcus pneumoniae*
  - tularemia) zoonosis

- urgent disease vs long term trends
- well-known vs newly introduced
- different modes of transmission

Capture-Recapture for Sensitivity

- method: capture-recapture

$$\widehat{x_{00}} = \frac{x_{10} * x_{01}}{x_{11}}$$

$$\frac{4.1}{X_{00}} = \frac{24 * 42}{245}$$

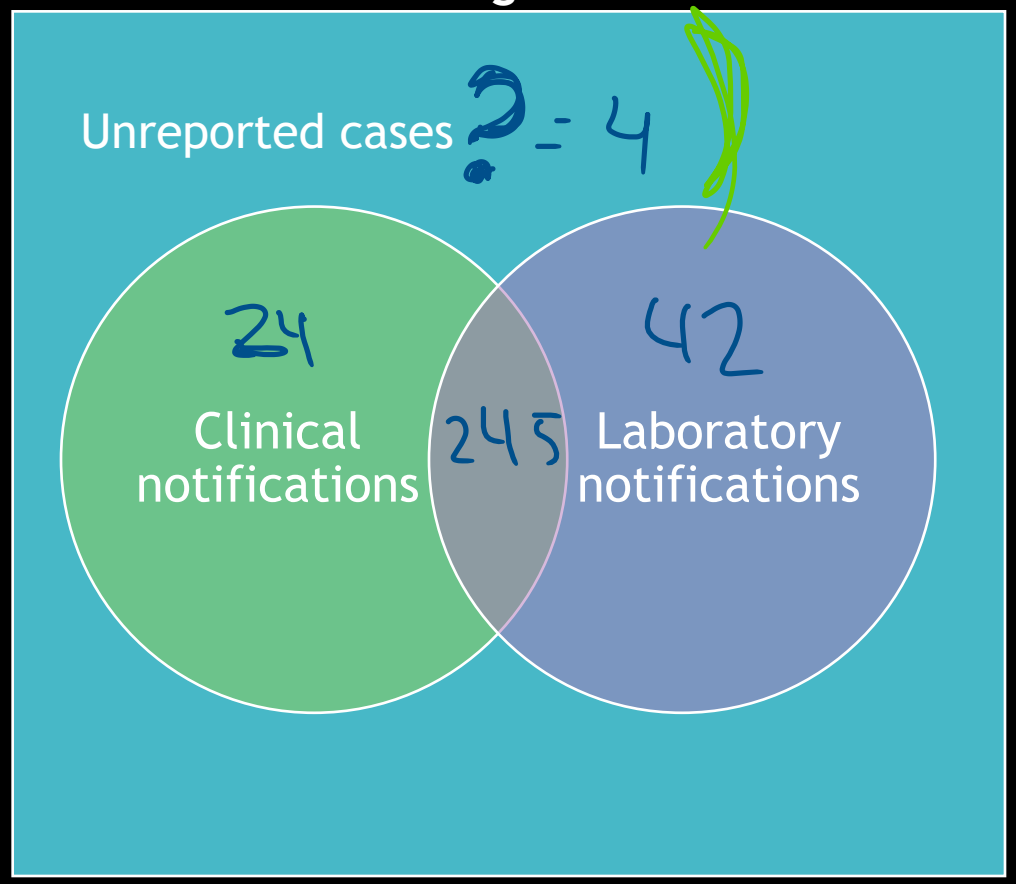
- calculate sensitivity

clinic $= \frac{24 + 245}{315} =$

lab:

$$sens = \frac{reported\ cases}{total\ cases}$$

overall $= 24 + 245 + 42 = \frac{311 \text{ reported}}{315 \text{ total}} = 91.1\%$



Capture-Recapture for Sensitivity

- reports by source, combined, and estimation of total cases

Table 1. *Number of notifications and total number of cases as estimated by the capture–recapture method for each studied disease, 1998–2002*

Disease	Clinical notifications	Laboratory notifications	Overall notified cases	Estimated total cases
Meningococcal infection	287	269	311	315
Salmonellosis	22 236	22 739	23 185	23 202
PRP*	2187	3051	3410	3696
Tularaemia	817	646	855	890
Total	25 527	26 705	27 761	28 103

* Infection due to penicillin-resistant pneumococci.

reported + unreported

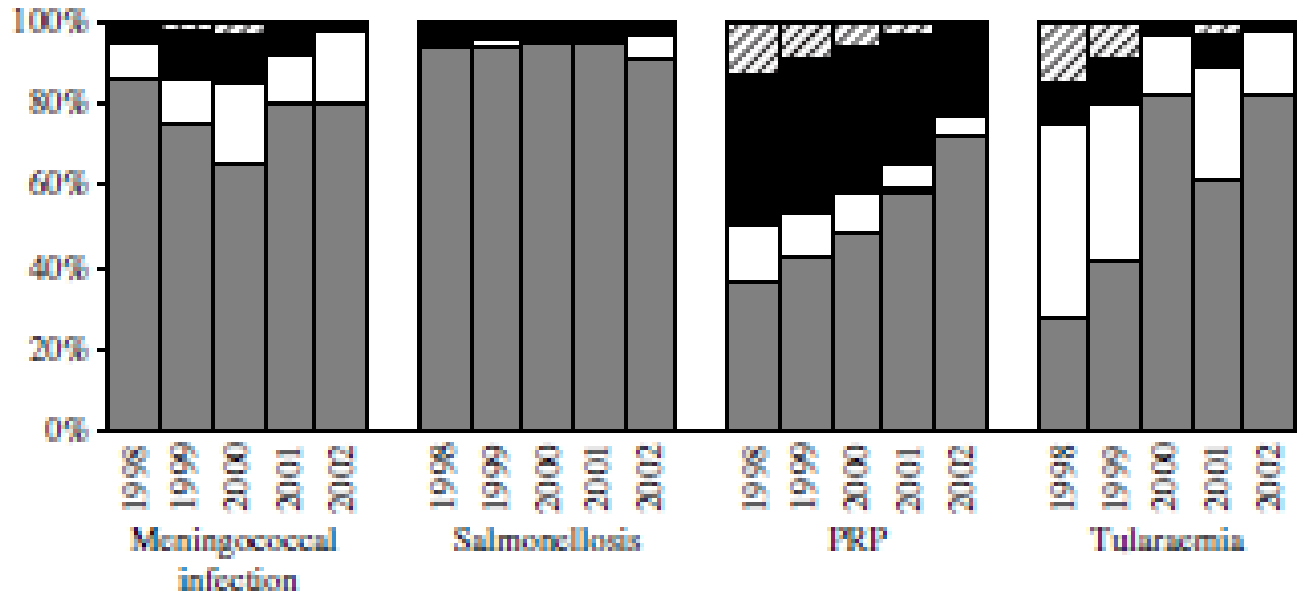


Fig. 2. Estimated sensitivity of the surveillance system for each disease and year, using the capture–recapture method. ■, Clinical and laboratory notifications; □, clinical notifications only; ■, laboratory notifications only; ▨, not reported.

Capture-Recapture for Sensitivity

- sensitivity by year for each disease
- also reported as overall sensitivity, clinical sensitivity, and laboratory sensitivity

- sensitivity affected by electronic reporting?

electronic not

	SmiNet-connected counties (n = 14672)*	Not SmiNet-connected counties (n = 13192)*
Clinical and laboratory notification	13 270 (90·4 %)	11 201 (84·9 %)
Clinical notification only	334 (2·3 %)	722 (5·5 %)
Laboratory notification only	1042 (7·1 %)	1192 (9·0 %)
Neither clinical nor laboratory notification†	26 (0·2 %)	77 (0·6 %)
Sensitivity of clinical notifications	92·7 %	90·4 %
Sensitivity of laboratory notifications	97·5 %	93·9 %

* Counties connected during the study period have contributed to respective groups by year of study.
† Estimated from the capture–recapture method with 95 % confidence intervals.

higher

Capture-Recapture for Sensitivity

- comparison of source sensitivity based on which sources were connected to the national electronic surveillance system

- making changes → impact

Workshop Schedule

Time	Topics
~~2:00—2:15 pm~~	~~Surveillance Attributes~~
~~2:15—2:45 pm~~	~~System Assessment~~
~~2:45—3:00 pm~~	~~Example Assessment~~
3:00—3:10 pm	Break
3:10—4:00 pm	Mapping

Geography in Medicine & Public Health

- Widespread use of GIS, mapping, spatial analyses in medical fields and public health
- Basic tenet of epidemiology: analysis by person, **place** and time

Cholera outbreak London John Snow

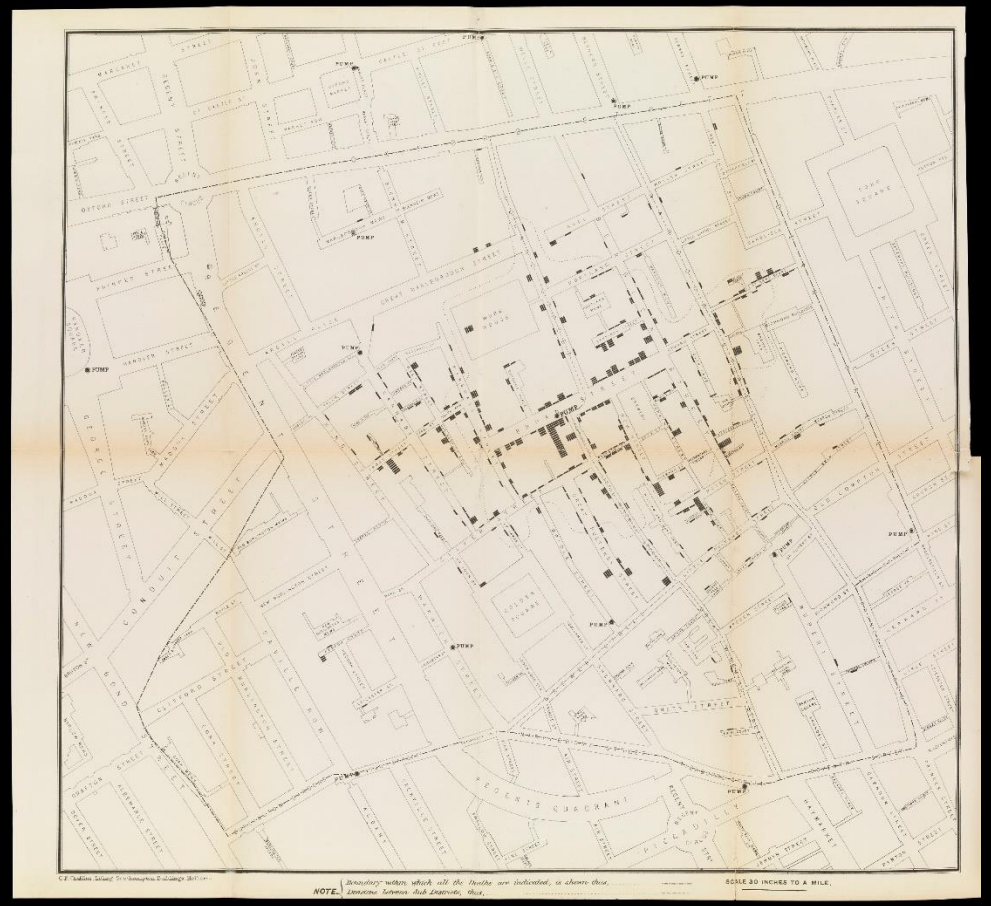
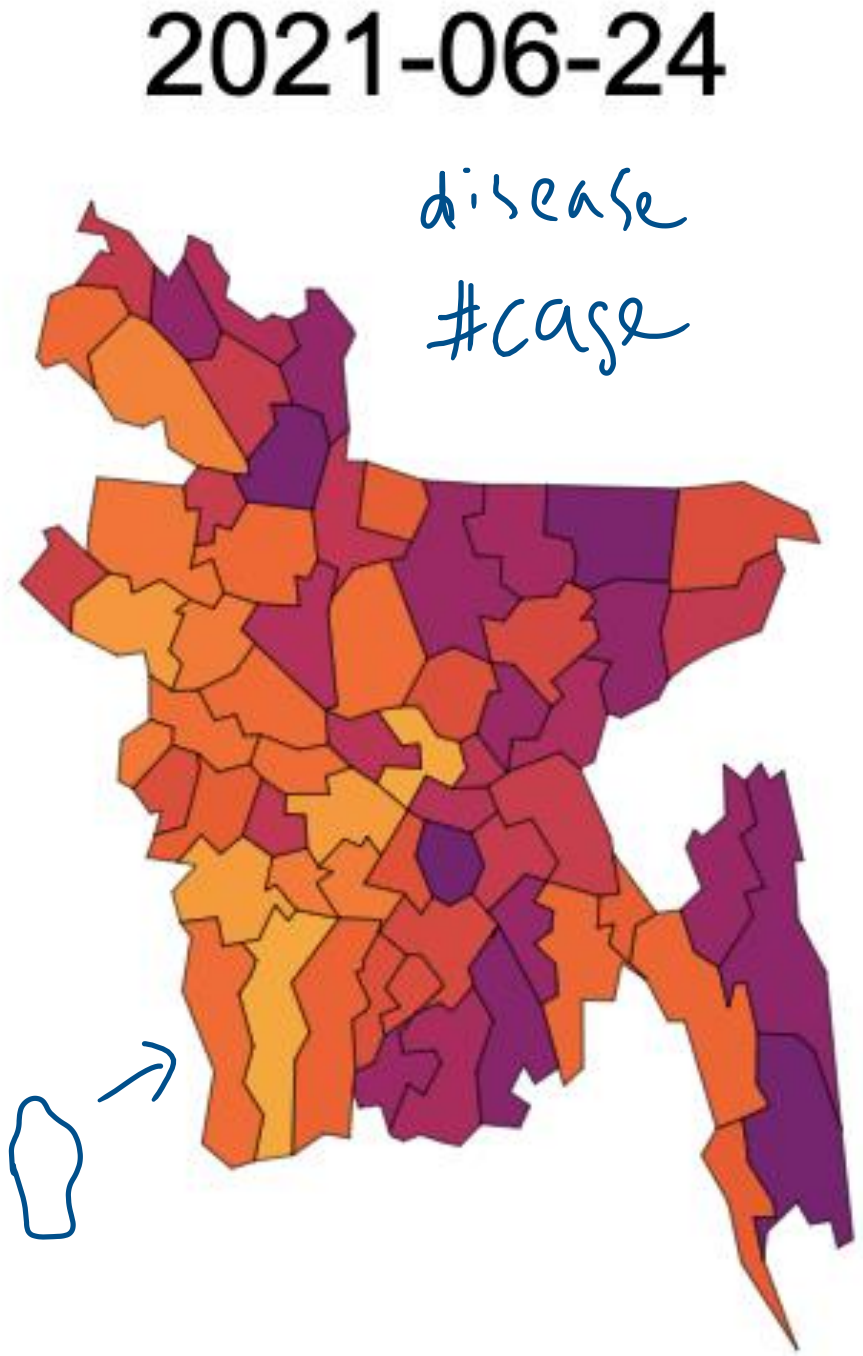


Image credit: Wikimedia.org

Applications in Public Health

Geography Method	Example Use
Visualization/mapping	Understanding distribution, health communication
Spatial exploration	Source/risk proximity, case definitions
Cluster	Hotspots/outbreaks
Spatial modeling	Risk prediction, dispersion/exposure modeling, spatially-adjusted models

Smith *et al.*, Euro Surveill.



Mapping Data

- Many types of maps!
- spatial visualization of data
  - choropleth map
    - color in each administrative area corresponds to our data of interest
    - areal data (2D, polygons)
    - popular visualization

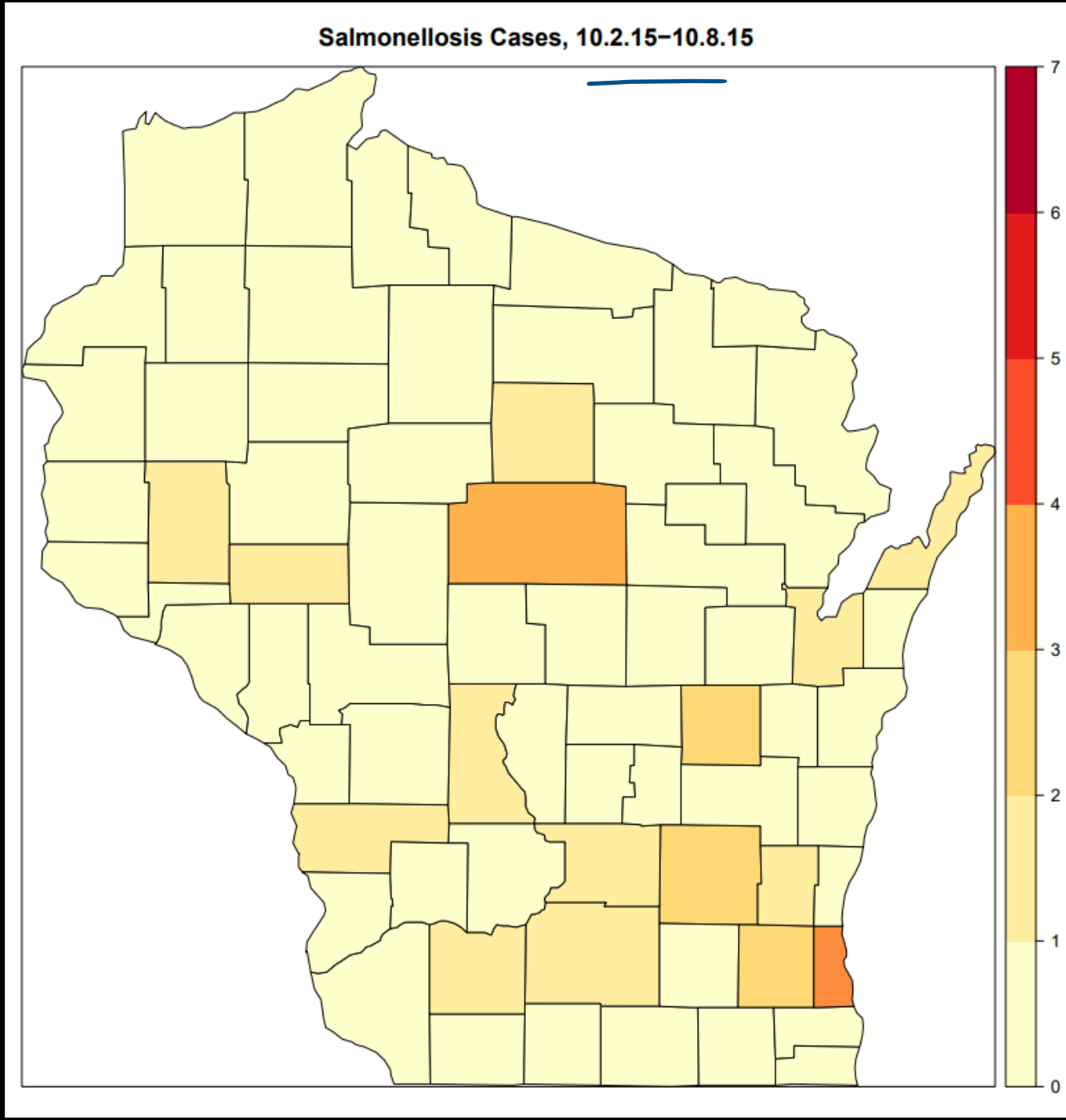

Salmonellosis Cases, 10.2.15−10.8.15
7
6
5
4
3
2
1
0

Figure 3 Maps of Smoothed Standardized Infection Ratio for Identified Clusters

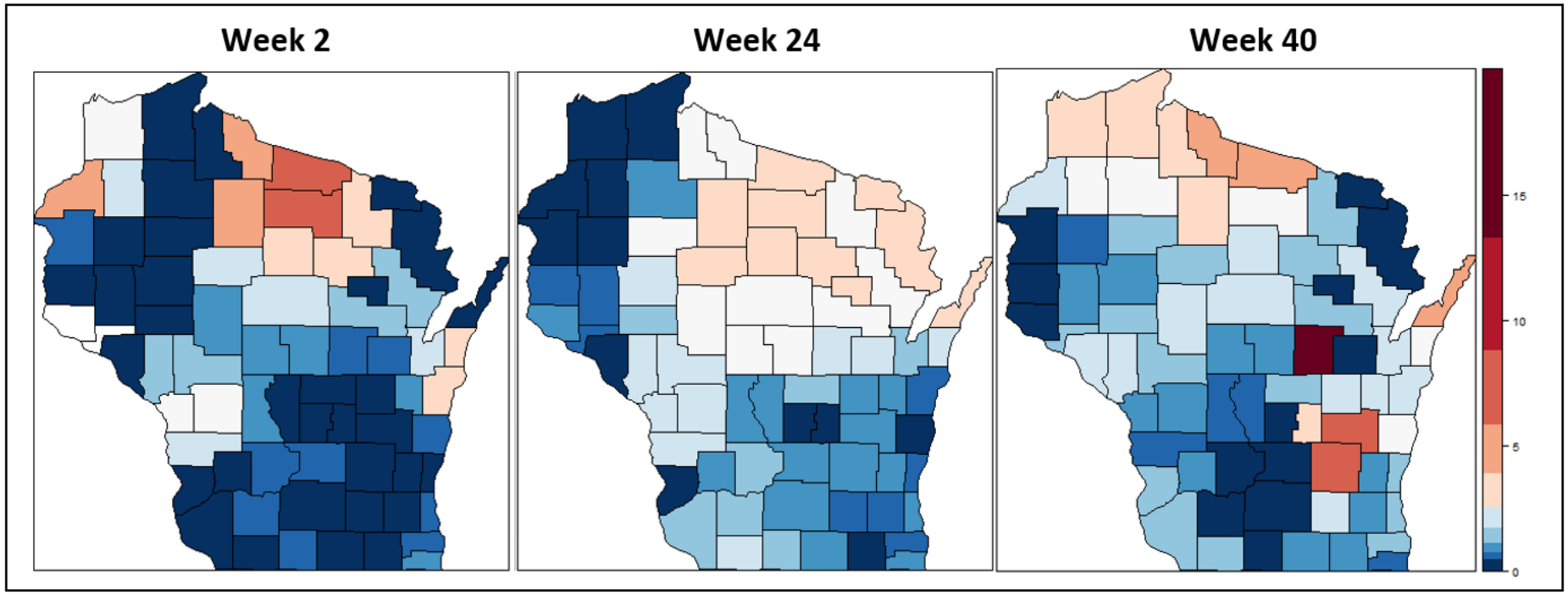
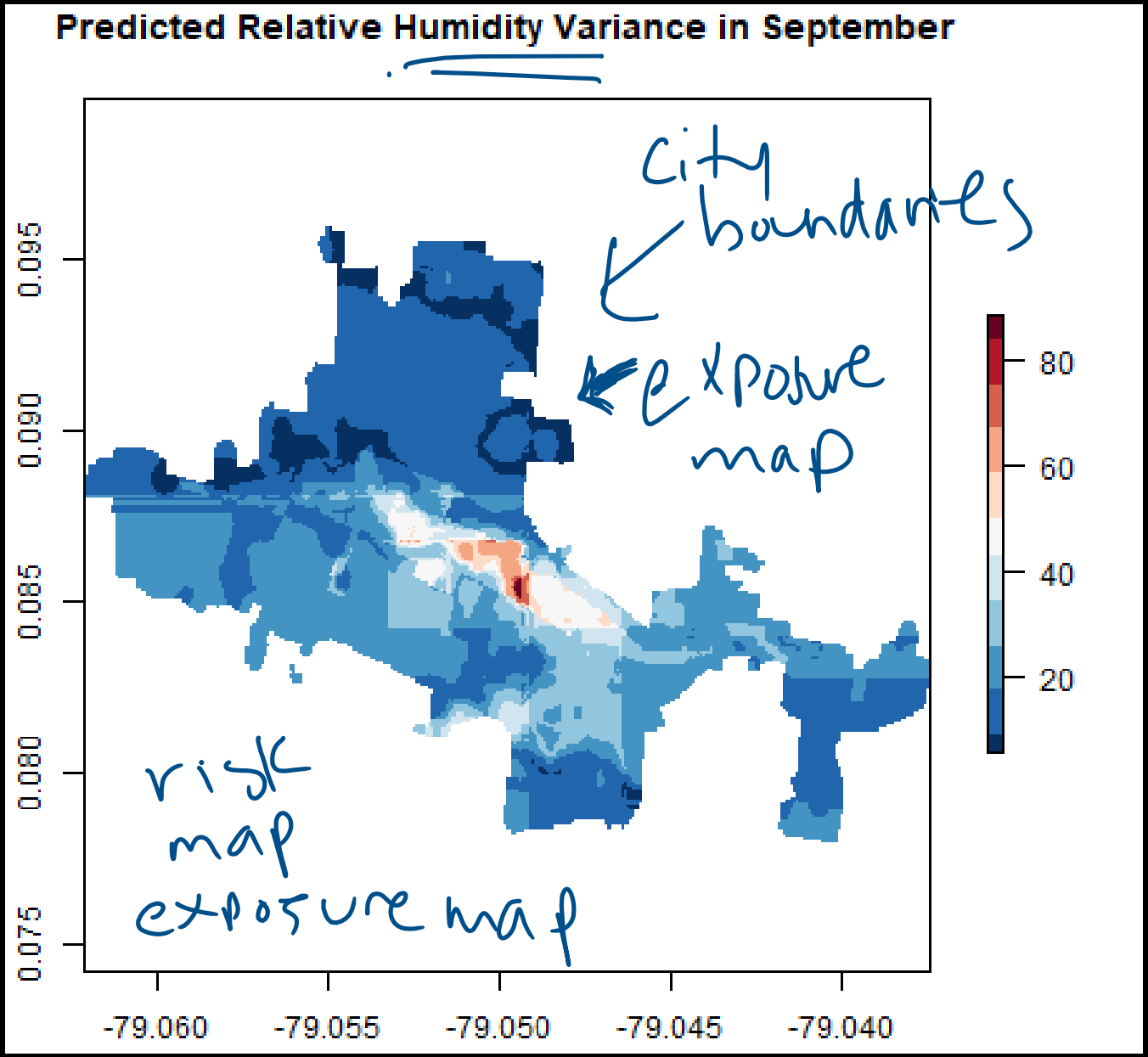


Table 2: Cluster Location Kulldorff's spatial scan statistic calculated for weeks with significant clustering. Clusters listed are most likely clusters.

Week	Kulldorff p-value	Cluster Size	Cluster Location
2	0.048	3	Oneida, Vilas, Lincoln
24	0.036	12	Vilas, Oneida, Lincoln, Langlade, Forest, Price, Marathon, Iron, Menominee, Taylor, Florence, Sawyer
31	0.381	N/A	N/A
40	0.001	3	Dodge, Fond du Lac, Jefferson
44	0.989	N/A	N/A


Predicted Relative Humidity Variance in September
0.095
0.090
0.085
0.080
0.075
-79.060
-79.055
-79.050
-79.045
-79.040
80
60
40
20
city boundaries
exposure map
risk map
exposure map

Software & Tools

more comprehensive

$$ → • ArcGIS mapping + analysis

$$

free → • QGIS mapping + analysis

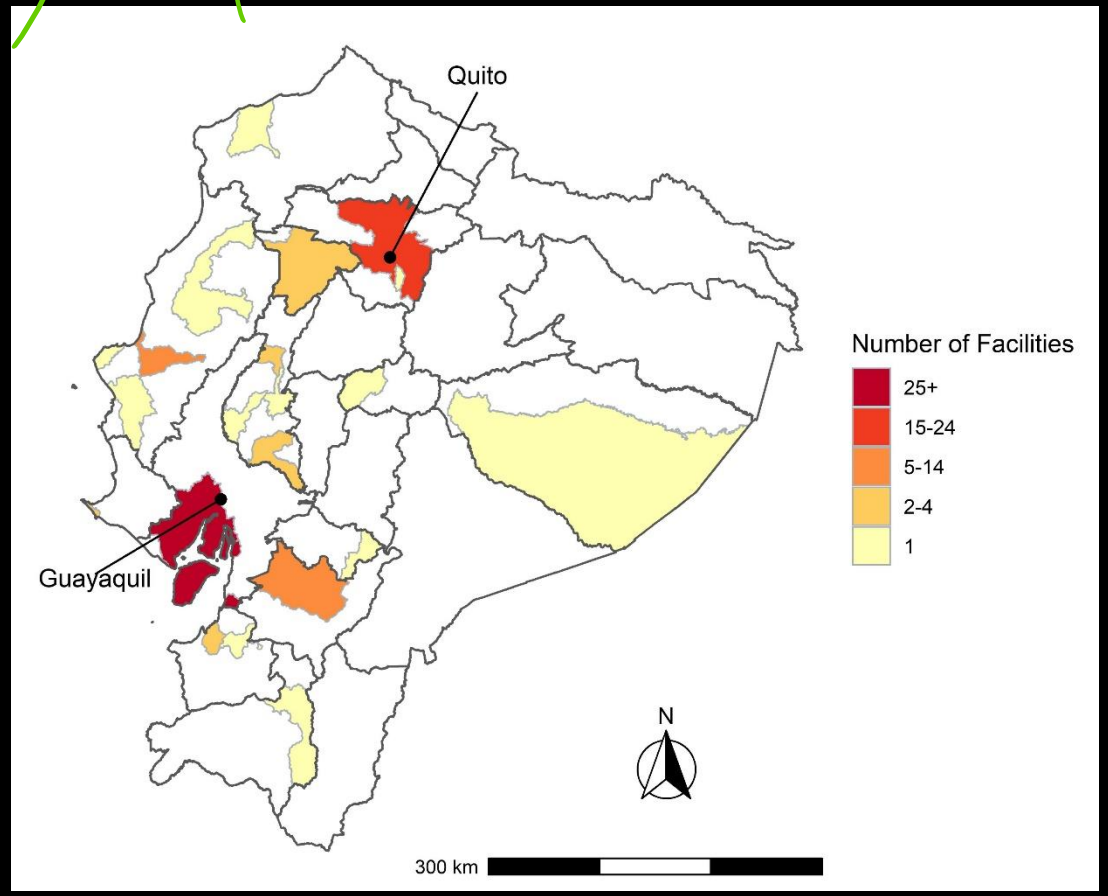
free → • GeoDa – analysis

free → • GoogleMaps/Google Earth ⇒ maps

$$ • Tableau → mapping

free → • SaTScan → analysis

free → • R mapping + analysis

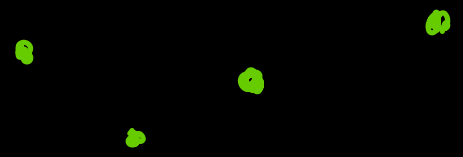


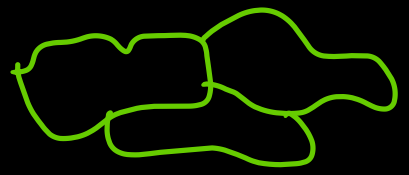
How Do We Store Locational Data?

- Lat/Lon pairs coordinates, point
- Addresses number, streetname) rural :(
- Place names
- Driving routes (turn by turn directions)
- Boundaries of areas
- By state/county/region/etc.

What Kinds of Objects?

[0-D] Single point locations	[1-D] Line segments	[2-D] Polygons
• (Latitude, Longitude) • local coordinate systems	• streets • boundaries	• counties, states, nations • lakes • building footprints



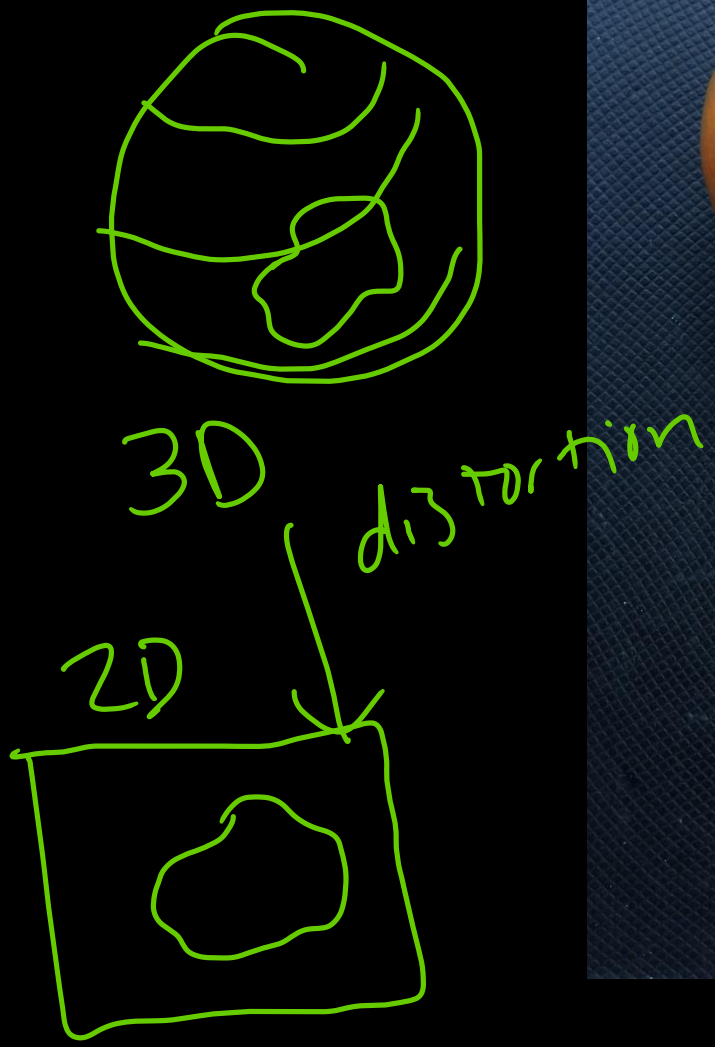


How Do We Present Location Data?

- Points/icons on a map
- Paths of roads on a map
- Images of regions on a map

- ON A MAP (2D image of our 3D planet)

Map Making is Challenging



Map Making is Challenging

- Impossible to avoid distortion on 2D image
- Geographic coordinates / projected coordinates

Google maps

[WGS84, NAD83]

[Mercator, MO-east]

[lat/long]

[meters, feet, etc.]

-90.4, 78.2

"Not only is it easy to lie with maps, it's *essential*."

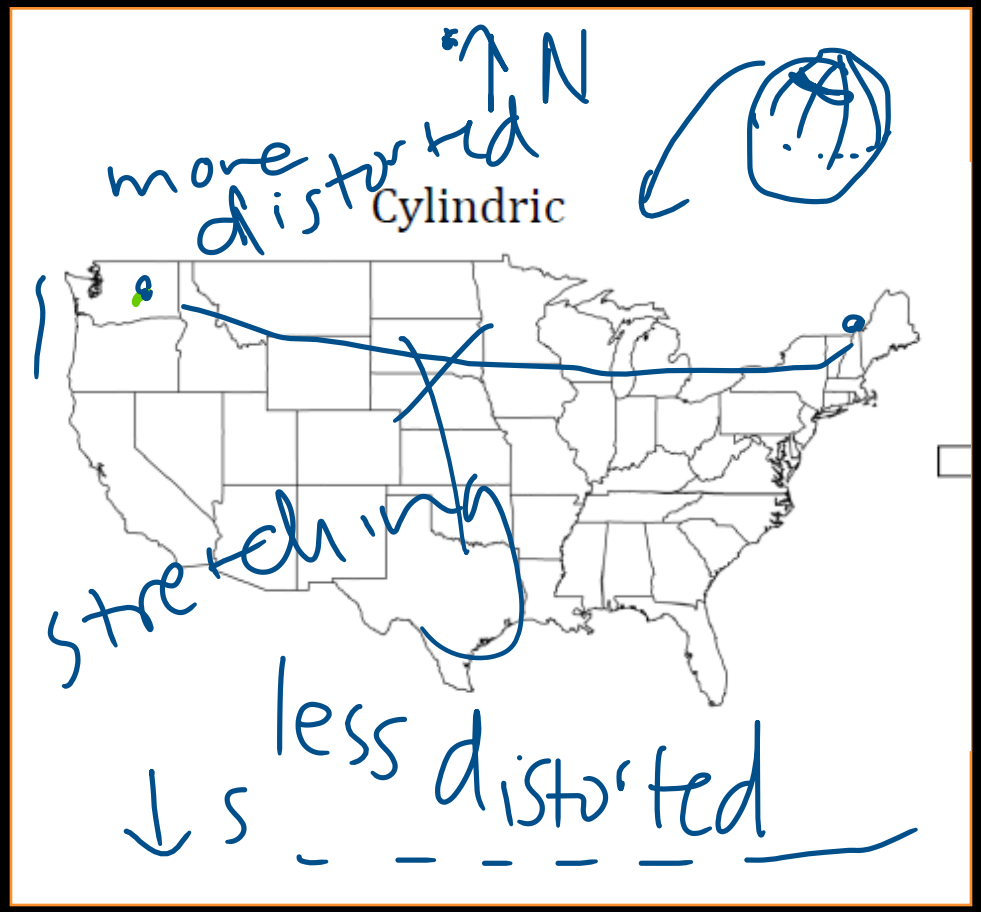
- Mark Monmonier, *How to Lie with Maps*

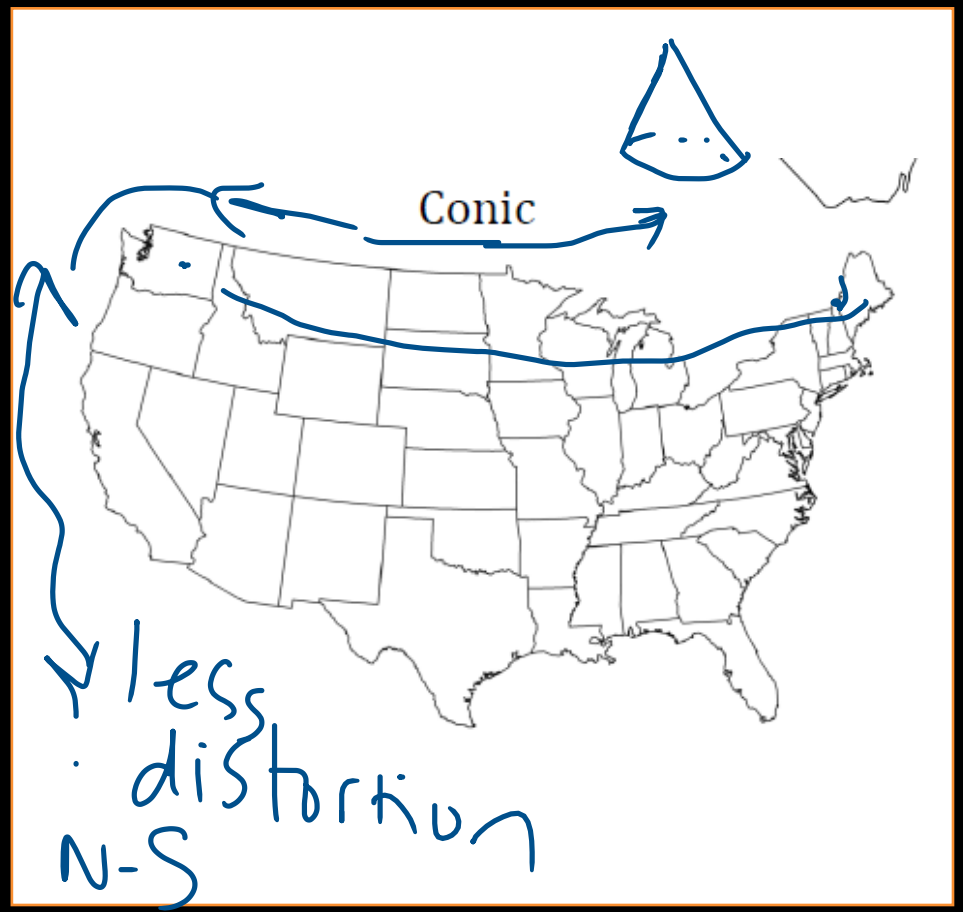
17S UTM → specific to location on Earth

Ecuador

small area
less distortion

Map Projections

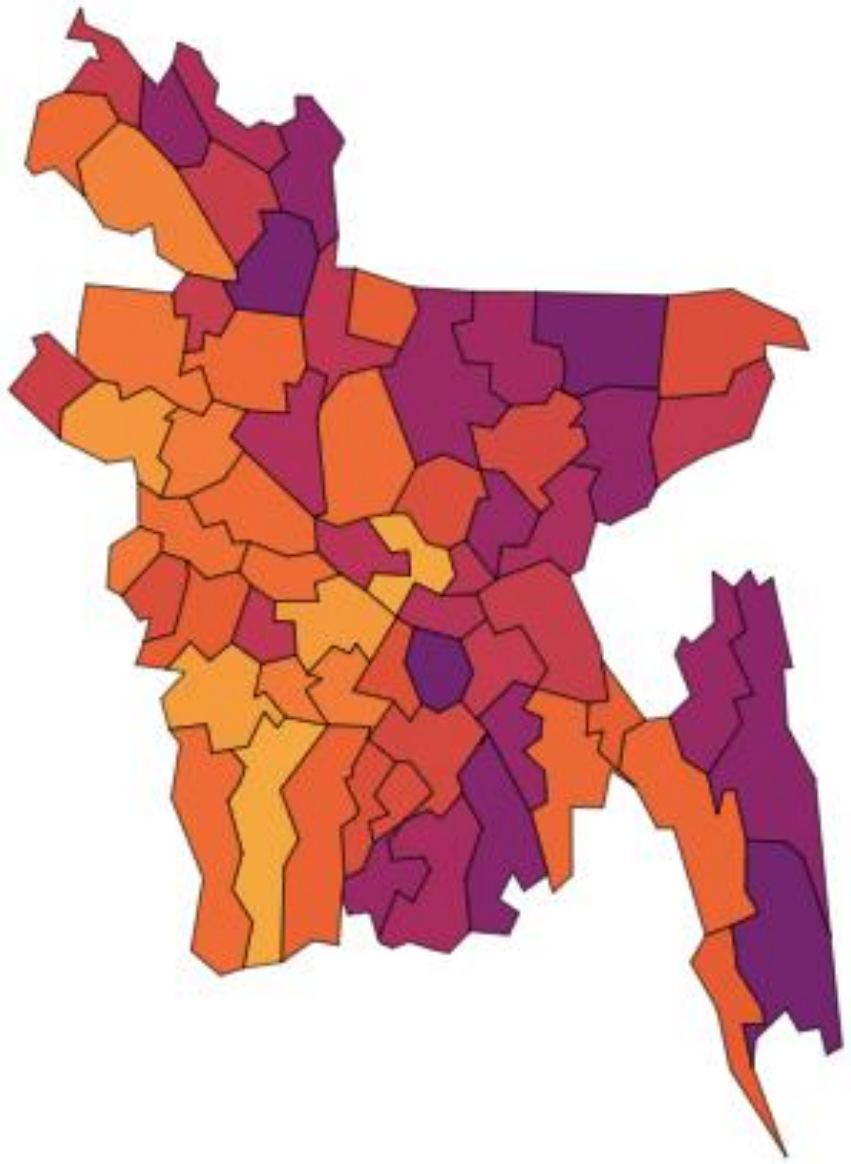




Some Spatial File Types

- Shapefile (the geospatial equivalent of “csv”)
- KML (works with Google Earth)
- GeoJSON
- Geodatabase (ESRI proprietary format)
- SpatialPolygonsDataFrame (inside R)

2021-06-24



Spatial & Mapping Packages

- ggmap: use ggplot2 for mapping
- rgdal: spatial data processing
- rgeos: vector processing
- maptools: mapping features
- sf: simplifies spatial data

ggplot2
- tidyverse
- data visuals

- RColorbrewer: color schemes for mapping
- viridis: additional color schemes

- many, many more!

COLORS IN DATA VISUALIZATION

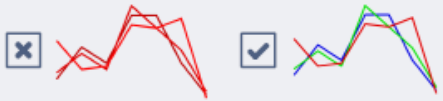
RULES

INTUITIVENESS

Use intuitive colors. When choosing them, consider what associations do they evoke.
If possible, use colors that audience will associate with your data anyway.

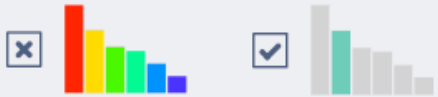
CLARITY

Use colors to make the data easier to read. Make sure your audience will be able to distinguish between the items shown in the visualization.



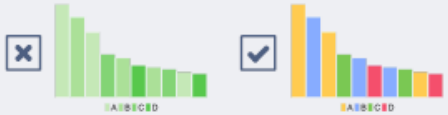
MODERATION

Use colors in moderation. For a simple dataset, a single color is preferable.
Use color as a strategic tool to highlight the important parts of your visual.



CLASSIFICATION

Don't use a gradient color palette for categories.
And the other way round - different colors for same measurement.

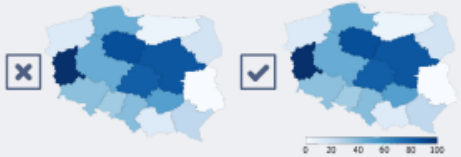


CONSISTENCY

Use colors consistently. Change colors if you want your audience to feel the change for the specific reason, but never simply for the sake of novelty.

EXPLAINABILITY

Make sure to explain to your audience what exactly used colors mean.
Remember to create a color key.



COLOR SCHEMES

Monochromatic - the simplest formula for harmony is monochromatic. Consists of different shades of one hue. Not a good choice if we want to highlight something.

Analogous - this scheme is composed of colors that are next to each other on the wheel. Usually they match up pretty well, making elegant and clear look.

Complementary - uses two colors which are opposite on the color wheel. With saturated colors makes very vibrant look. Try to tone down colors to avoid overvibrance, by adjusting saturation and lightness/darkness. Do not use with text with saturated colours.

Triadic - uses three colors that evenly spaced on the color wheel. Makes that none of colors is dominant and quite vibrant look.

Split-Complementary - variation of complementary scheme. Uses base color and two adjacent to its complementary color. Often this scheme is more pleasant to the eye than usual complementary scheme.

Tetradic - this scheme consists of four colors, two of them are complementary to other two. Choosing one color as dominant and the rest as accents, gives the best result.

Remember! Do not stick strictly to colors imposed by a scheme. These patterns are just starting points, you can create your own variations based on schemes above. Check also: paletton.com

PRO TIPS

The color grey is the most important color in data visualization.

The use of color should always be an intentional decision. Never let your tool make this important decision for you!

After creating you visualization, close your eyes and then look back at it, taking note of where your eyes are drawn first. Is it where you want your audience to focus?

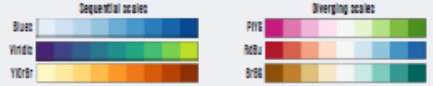
When picking colors consider the connotations colors have in other cultures.
You can check: informationisbeautiful.net/visualizations/colours-in-cultures

Remember about color deficiency issues (color blindness).
You can check: projects.susielu.com/viz-palette

COLOR PALETTES

QUANTITATIVE DATA - SEQUENTIAL OR DIVERGING COLORS

Color is used show variations in the data. The palette contains a sequence of colors that clearly indicate which values are larger or smaller than which other ones (sequential scale). It can also visualize the deviation of data values in one of two directions relative to a neutral midpoint (diverging scale). Diverging scale can be viewed as two merged sequential scales.

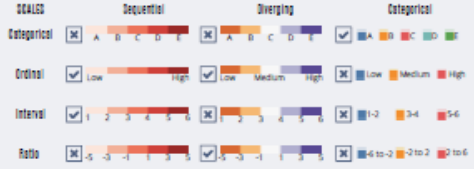


CATEGORICAL DATA - QUALITATIVE COLORS

Color is used to separate areas into distinct categories. The palette should consist of colors as distinct from one another as possible. The maximum number of categories that can be displayed is about 12 (practically speaking, probably fewer).

All examples are available in Seaborn library. Check also: medialab.github.io/iwanthue/

USAGE GUIDELINES



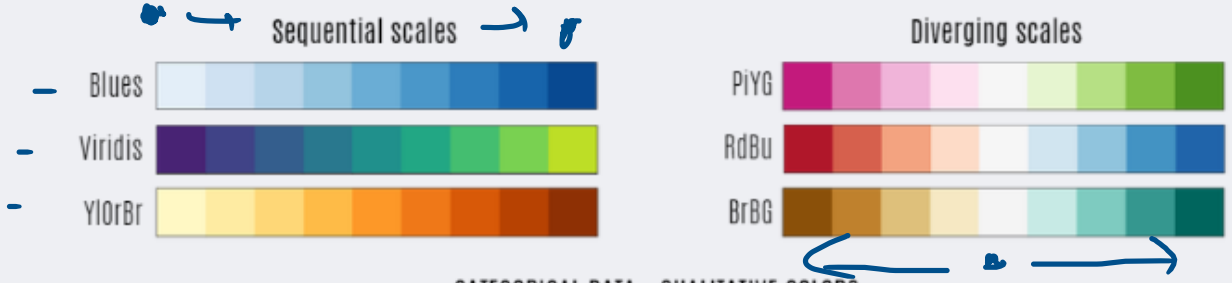
Authors:
Mateusz Krzyziński
Paweł Wojciechowski
Artur Żółkowski

Warsaw University of Technology
Faculty of Mathematics and Information Science

COLOR PALETTES

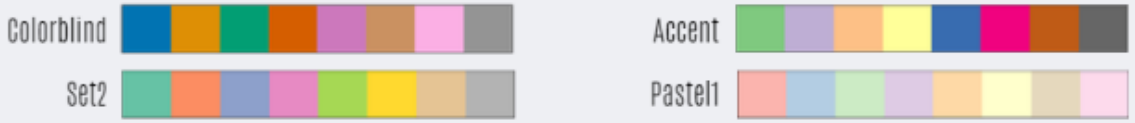
QUANTITATIVE DATA - SEQUENTIAL OR DIVERGING COLORS

Color is used show variations in the data. The palette contains a sequence of colors that clearly indicate which values are larger or smaller than which other ones (sequential scale). It can also visualize the deviation of data values in one of two directions relative to a neutral midpoint (diverging scale). Diverging scale can be viewed as two merged sequential scales.



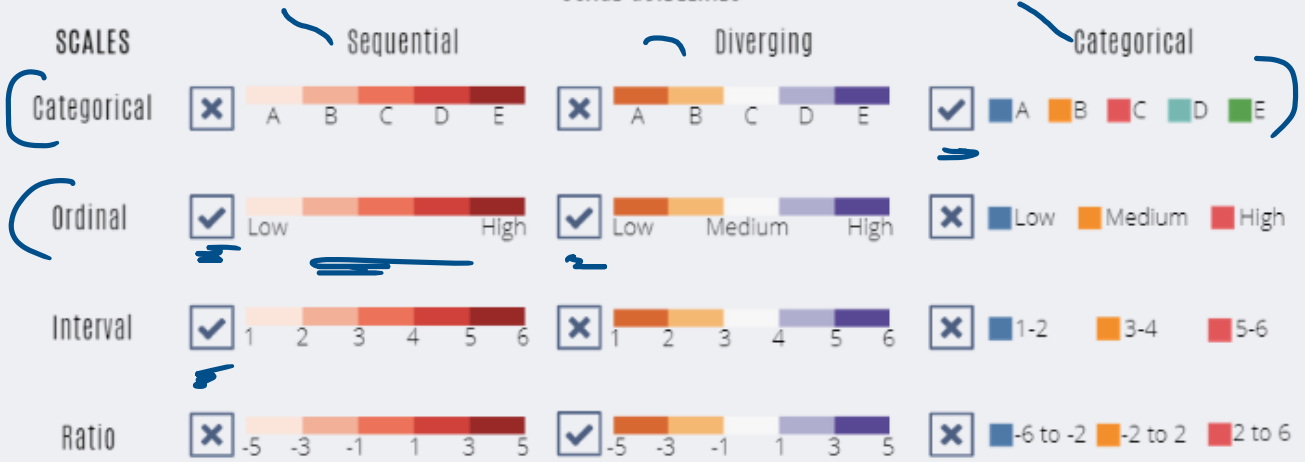
CATEGORICAL DATA - QUALITATIVE COLORS

Color is used to separate areas into distinct categories. The palette should consist of colors as distinct from one another as possible. The maximum number of categories that can be displayed is about 12 (practically speaking, probably fewer).

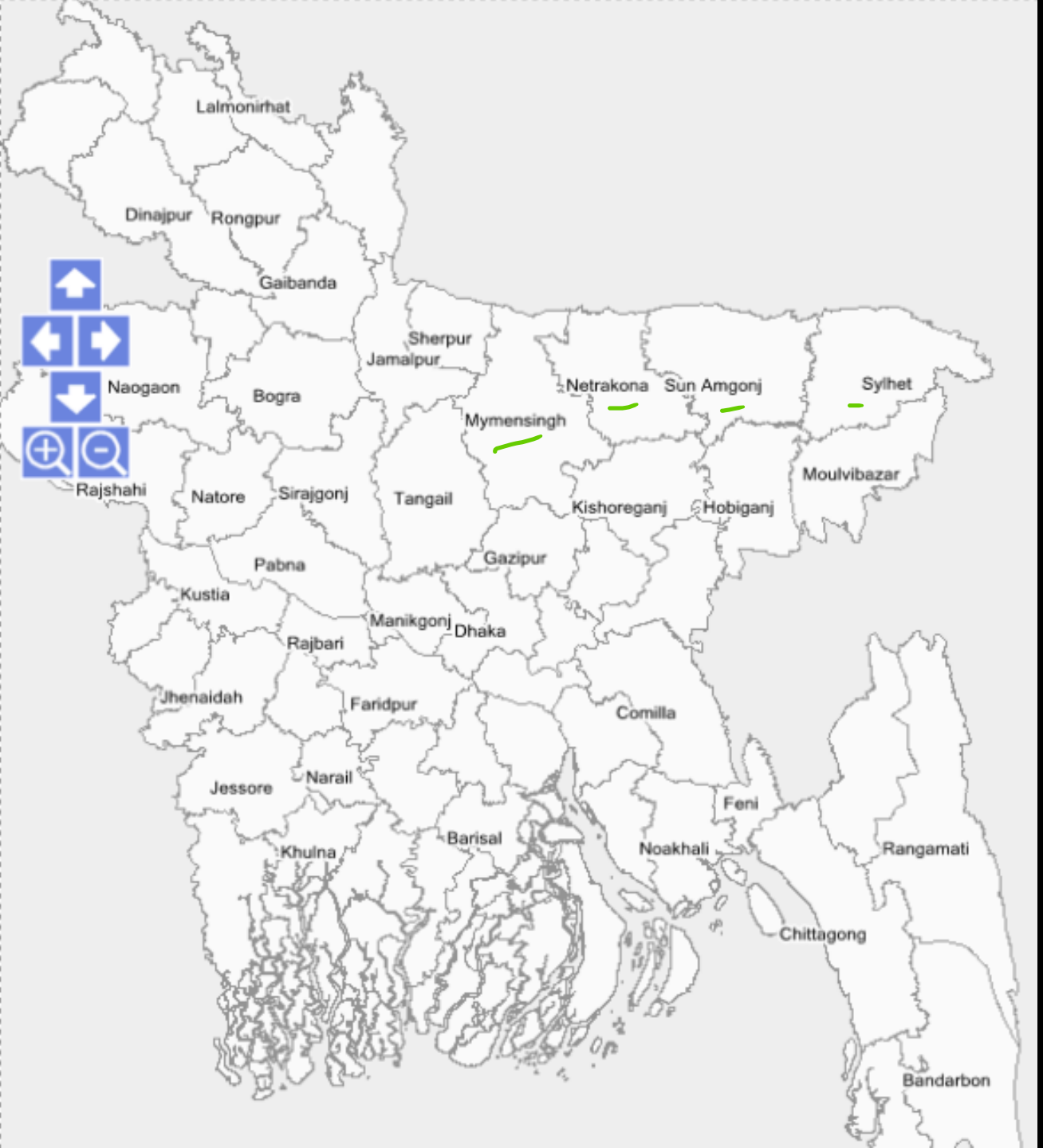


All examples are available in Seaborn library. Check also: medialab.github.io/iwanthue/

USAGE GUIDELINES



https://colorbrewer2.org/
Rcolorbrewer package

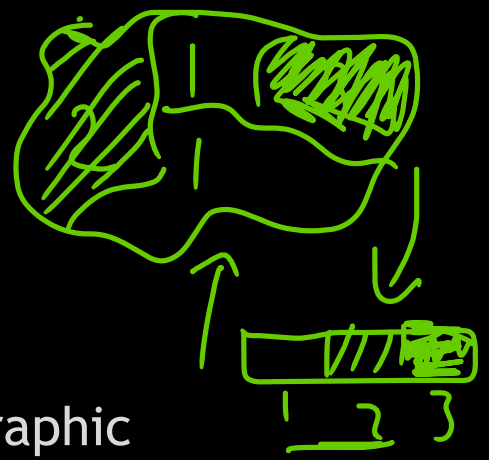


Mapping Data

shapefile

1. Spatial object: geographic representation of our area of interest
   - boundaries of administrative areas
   - where are they located? (CRS)
   - id/names for administrative areas
   - may be very large

Mapping Data



1. Spatial object: geographic representation of our area of interest
   - boundaries of administrative areas
   - where are they located? (CRS)
   - id/names for administrative areas
   - may be very large
2. Data to put on the map!
   - associated with each administrative area

Spatial Object

- Shapefile (the geospatial equivalent of “csv”)
- KMZ or KML: from GoogleEarth/GoogleMaps
- GeoJSON
- Geodatabase (ESRI proprietary format)
- SpatialPolygonsDataFrame (inside R)
- and others!

Some send sit

- Create your own — GPS phones + Google maps
- Download from online
  - GADM →
  - Several levels of administrative areas

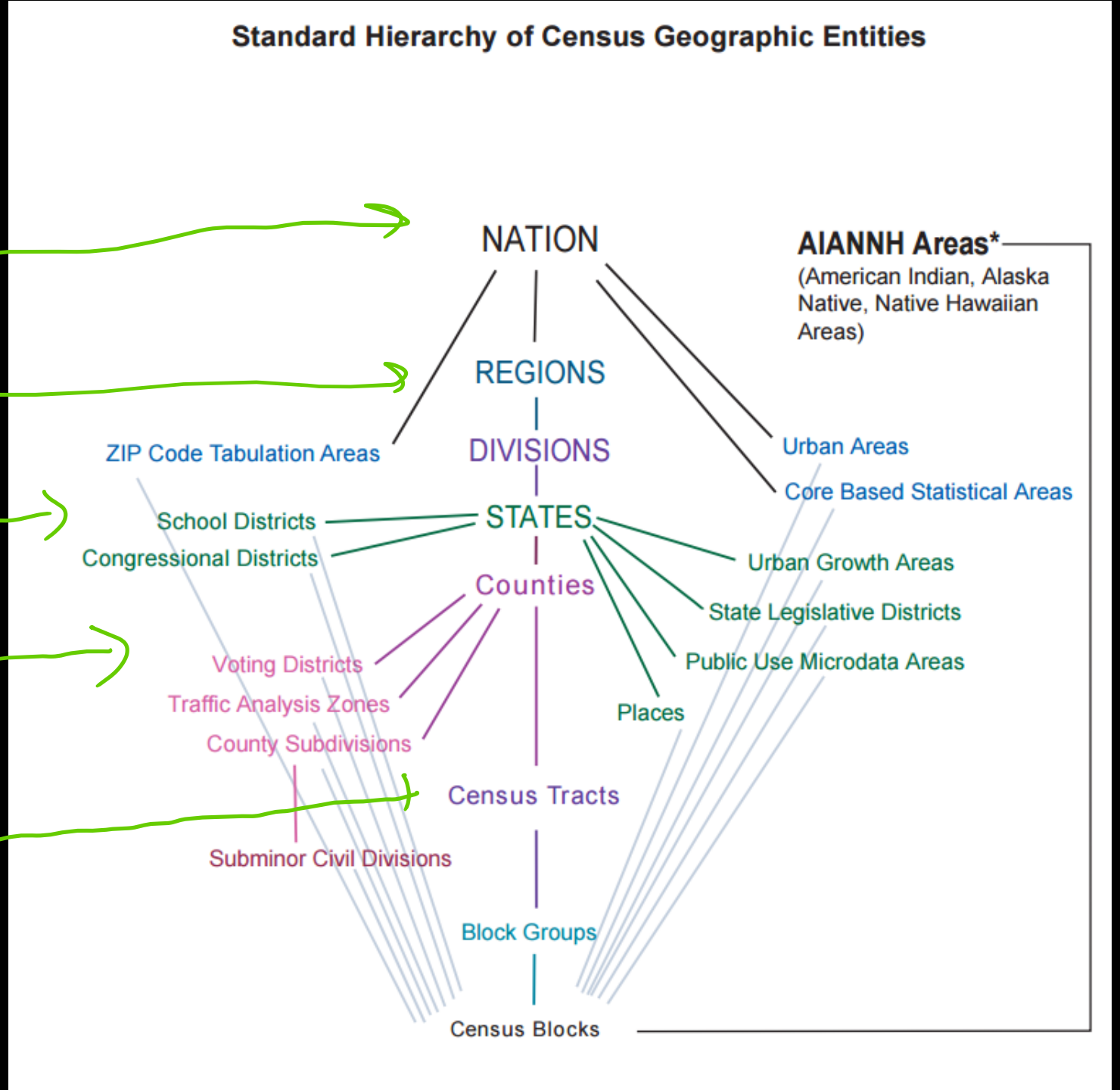
level 0

level 1

level 2

level 3

level 4



Spatial Object

shapefile + data

- administrative same name

- Most Important!!
  - Names of administrative areas must match your dataset!
  - This is how you will match your dataset to your spatial object
- Dataset gets merged to spatial object
  - when plotting, we assign a color for the area based on the value of the data variable you want to plot

+ other components

color

shape info
data