ETC5521: Diving Deeply into Data Exploration

Exploring data having a space and time context Part II

Professor Di Cook

Department of Econometrics and Business Statistics

You show me continents, I see the islands,
You count the centuries, I blink my eyes

~Björk

Outline

- Breaking up data by time, and by space
- Changing focus:
  - Maps of space over time
  - Exploring time over space with glyph maps
- Inference for spatial trends
- A flash back to the 1970s: Tukey's median polish
- Working with spatial polygon data
  - Making a choropleth map
  - Bending the choropleth into a cartogram
  - Tiling spatial regions

Approach

With spatiotemporaral data, you need to be able to pivot to focus on (1) time, (2) space, and (3) both together, although this latter task is harder.

The `cubble` object in the cubble R package, makes these operations easier.

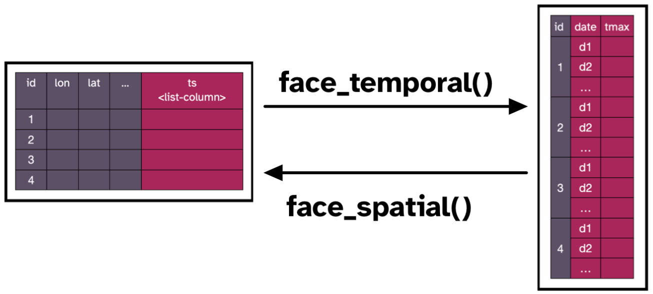
Example: Temperature change

6 years of monthly measurements of a 24x24 spatial grid from Central America collated by Paul Murrell, U. Auckland.

```
Rows: 41,472
Columns: 15
$ time        <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
$ id          <chr> "1-1", "1-2", "1-3", "1-4", "1-5", "1-…
$ lat         <dbl> -21, -21, -21, -21, -21, -21, -21, -21…
$ long        <dbl> -114, -111, -109, -106, -104, -101, -9…
$ elevation   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,…
$ date        <dttm> 1995-01-01, 1995-01-01, 1995-01-01, 1…
$ cloudlow    <dbl> 31, 32, 32, 39, 48, 50, 51, 52, 54, 56…
$ cloudmid    <dbl> 2.0, 2.5, 3.5, 4.0, 4.5, 2.5, 4.5, 5.0…
$ cloudhigh   <dbl> 0.5, 1.5, 1.5, 1.0, 0.5, 0.0, 0.0, 0.0…
$ ozone       <int> 260, 260, 260, 258, 258, 258, 256, 258…
$ pressure    <int> 1000, 1000, 1000, 1000, 1000, 1000, 10…
$ surftemp    <dbl> 297, 297, 297, 297, 296, 296, 296, 296…
$ temperature <dbl> 297, 296, 296, 296, 296, 295, 296, 295…
$ month       <ord> Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan…
$ year        <int> 1995, 1995, 1995, 1995, 1995, 1995, 19…
```

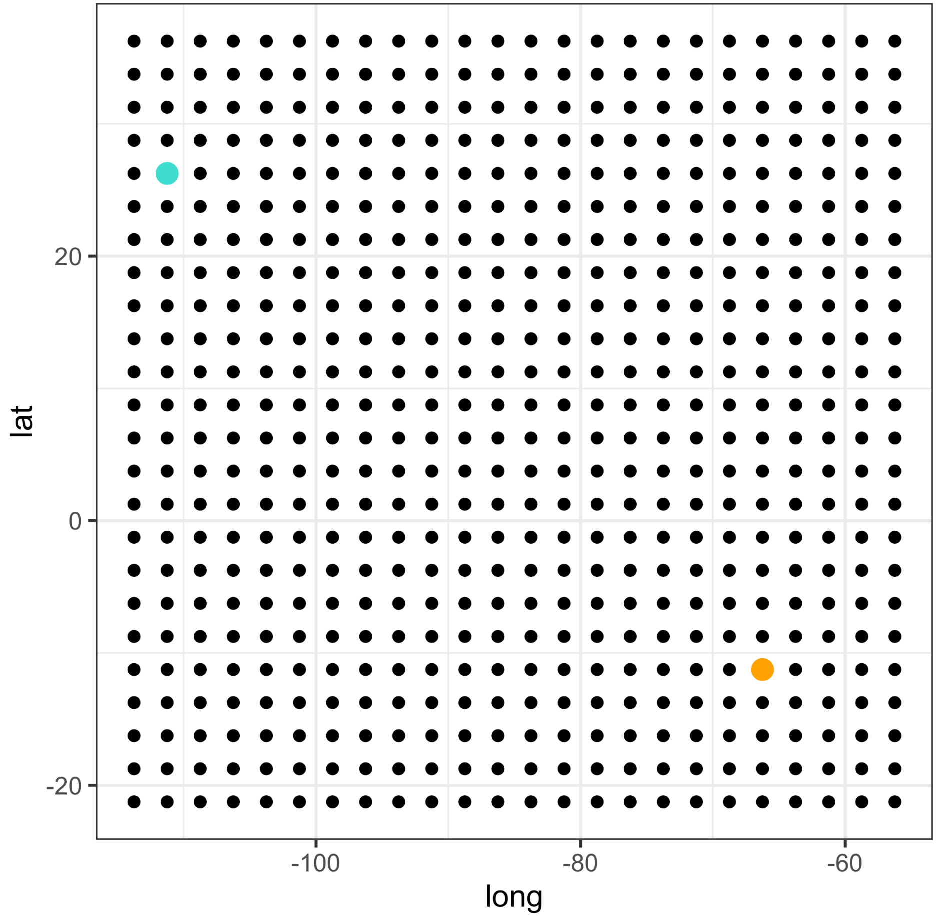
▶ Code

```
# cubble:   key: id [576], index: time, nested form
# spatial:  [-113.75, -21.25, -56.25, 36.25], Missing CRS!
# temporal: time [int], date [dttm], cloudlow [dbl],
#   cloudmid [dbl], cloudhigh [dbl], ozone [int], pressure
#   [int], surftemp [dbl], temperature [dbl], month [ord],
#   year [int]
   id      lat   long elevation ts
   <chr> <dbl>  <dbl>     <dbl> <list>
 1 1-1   -21.2 -114.          0 <tibble [72 × 11]>
 2 1-2   -21.2 -111.          0 <tibble [72 × 11]>
 3 1-3   -21.2 -109.          0 <tibble [72 × 11]>
 4 1-4   -21.2 -106.          0 <tibble [72 × 11]>
 5 1-5   -21.2 -104.          0 <tibble [72 × 11]>
 6 1-6   -21.2 -101.          0 <tibble [72 × 11]>
 7 1-7   -21.2  -98.8         0 <tibble [72 × 11]>
 8 1-8   -21.2  -96.2         0 <tibble [72 × 11]>
 9 1-9   -21.2  -93.8         0 <tibble [72 × 11]>
10 1-10  -21.2  -91.2         0 <tibble [72 × 11]>
```



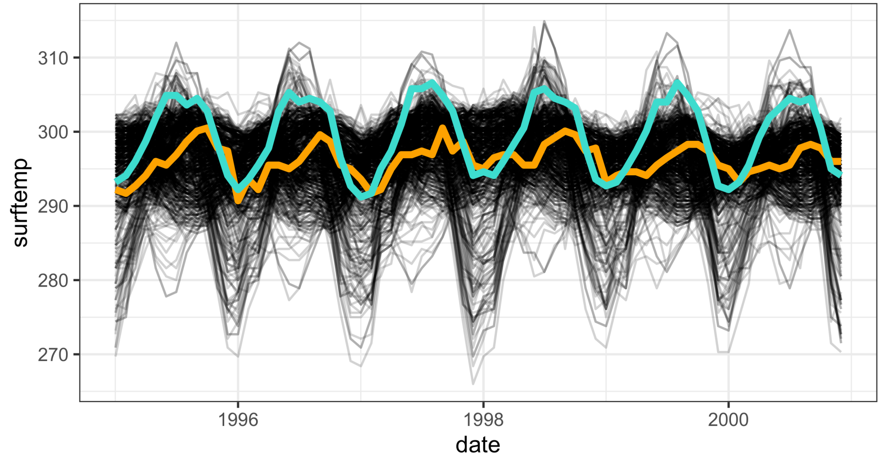
Spatial and temporal

▶ Code



20
0
-20
lat
-100
-80
-60
long

▶ Code



310
300
290
280
270
surftemp
1996
1998
2000
date

Pre-processing of time and space

> Think of time and space as ordered categorical variables.

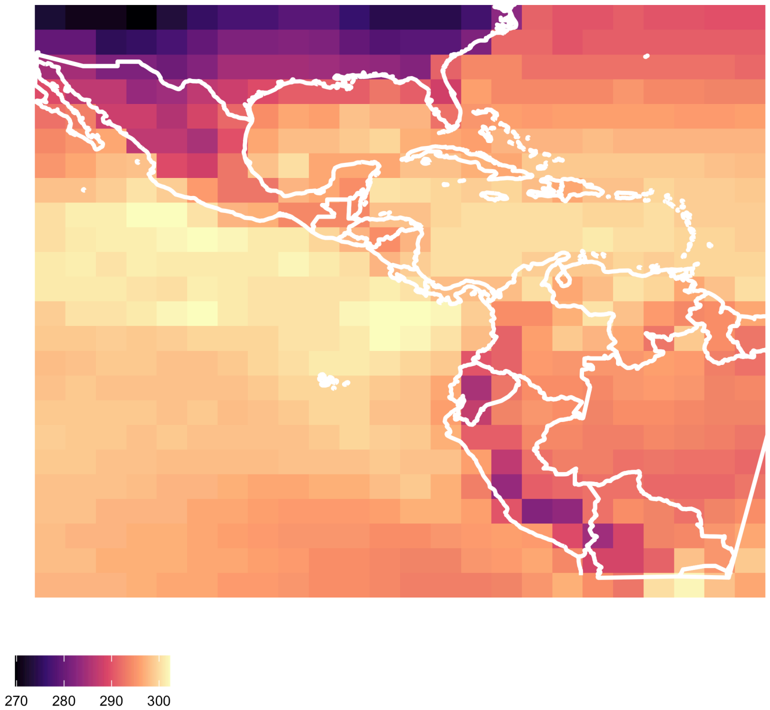
- Time may need to be converted to categories.
- Spatial variable *might* need to be discretised, or gridded.

This data is already gridded. Time is an integer from 1 to 72 (6 years of 12 months), as well as a date, and month and year. Space is a 24x24 grid of longitude and latitude, and also provided as an integer 1 to 24 in both x and y.

Focus on spatial (1/2)

🖼️ Learn Code

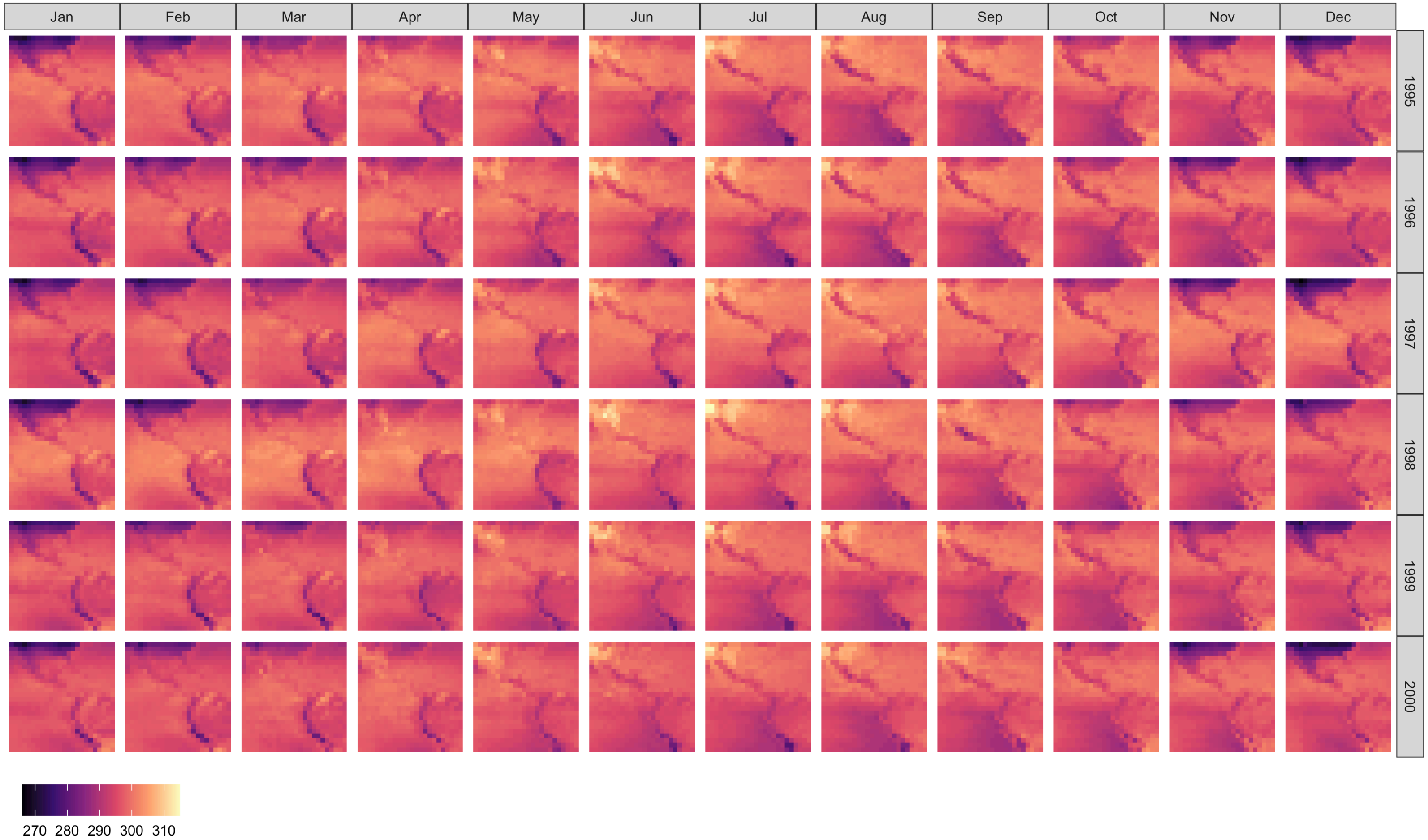
January 1995



Slice one month, and show gridded temperatures as a tiled display on spatial coordinates.

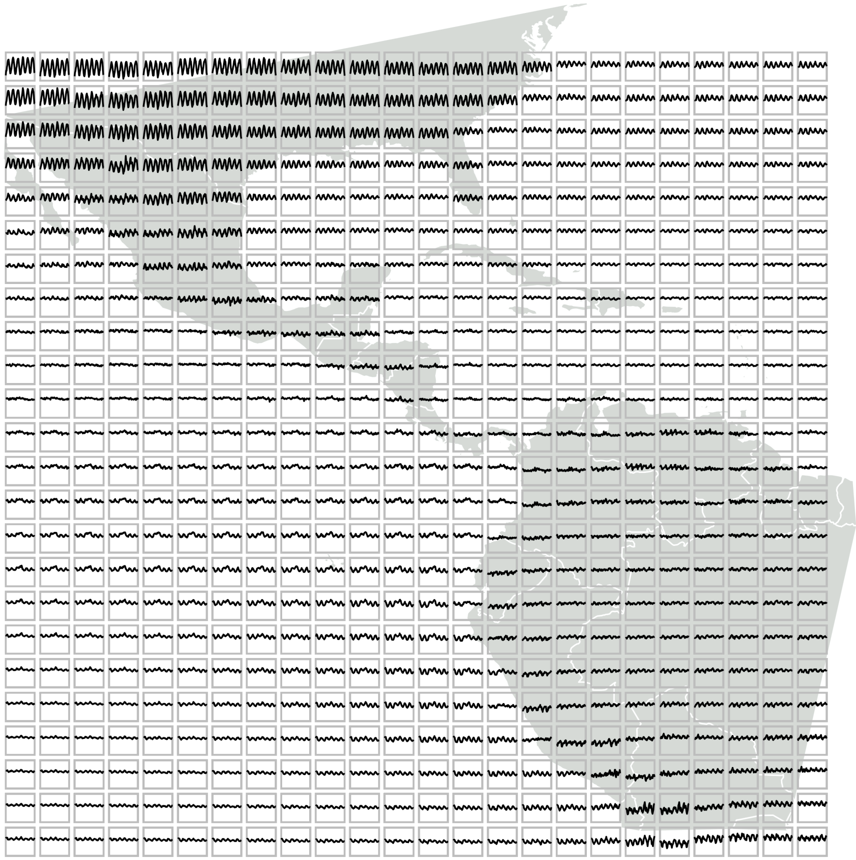
Focus on spatial (2/2)

🖼️ Learn Code


Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
1995
1996
1997
1998
1999
2000
270 280 290 300 310

Focus on temporal (1/4)

▶ Code



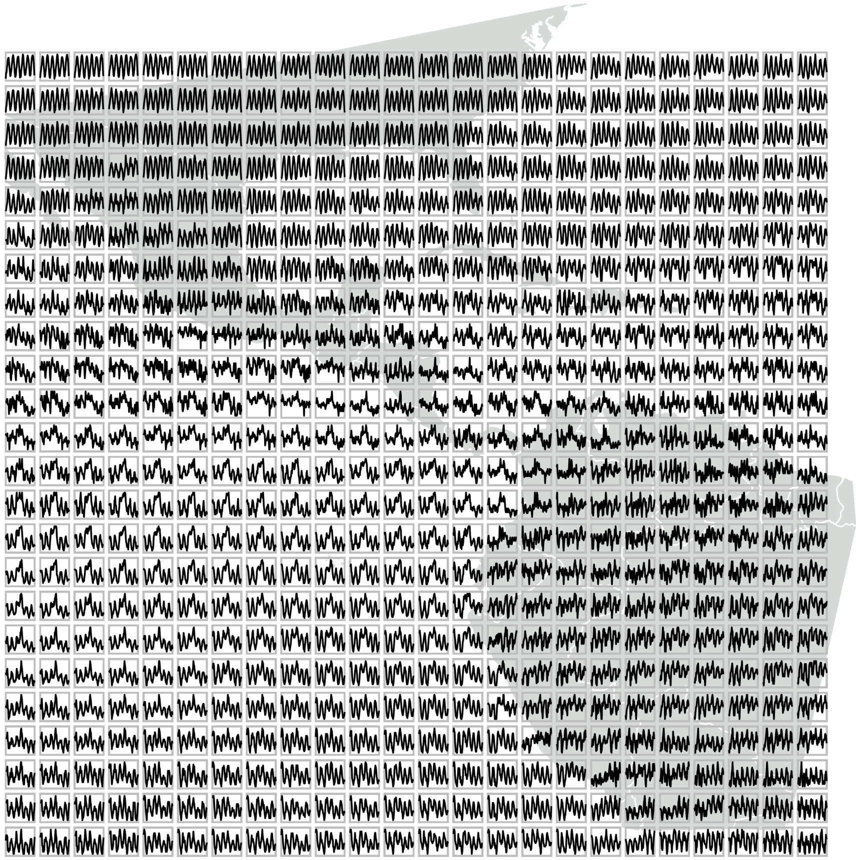
> A **glyphmap** shows a (small) time series at each spatial location.

- Several scaling choices:
  - global: overall min and max used to scale all locations
  - local: each location scaled on it’s own min/max

- Global scale used here
- Can see differences in the overall magnitude, particularly north to south.

Focus on temporal (2/4)

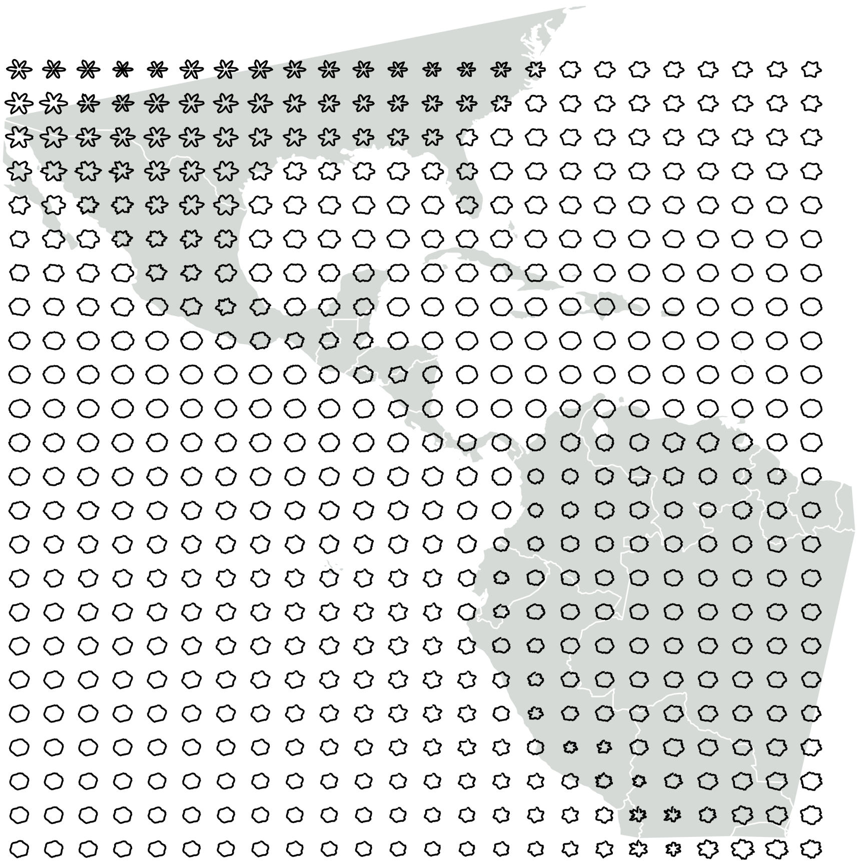
▶ Code



- Note: Local scale used, min/max for each spatial location
- El Nino year in equatorial region may be visible.
- Notice also odd patterns on the west (Andes mountains) of South America.

Focus on temporal (3/4)

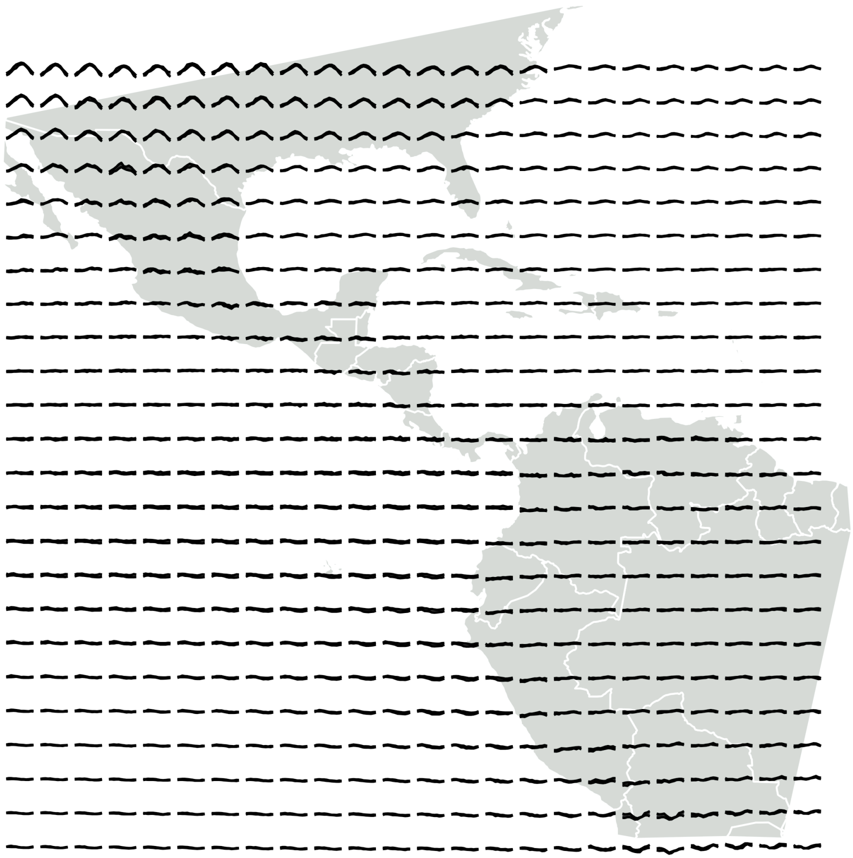
▶ Code



- Note: Local scale used, min/max for each spatial location, and polar coordinates used

Focus on temporal (4/4)

▶ Code

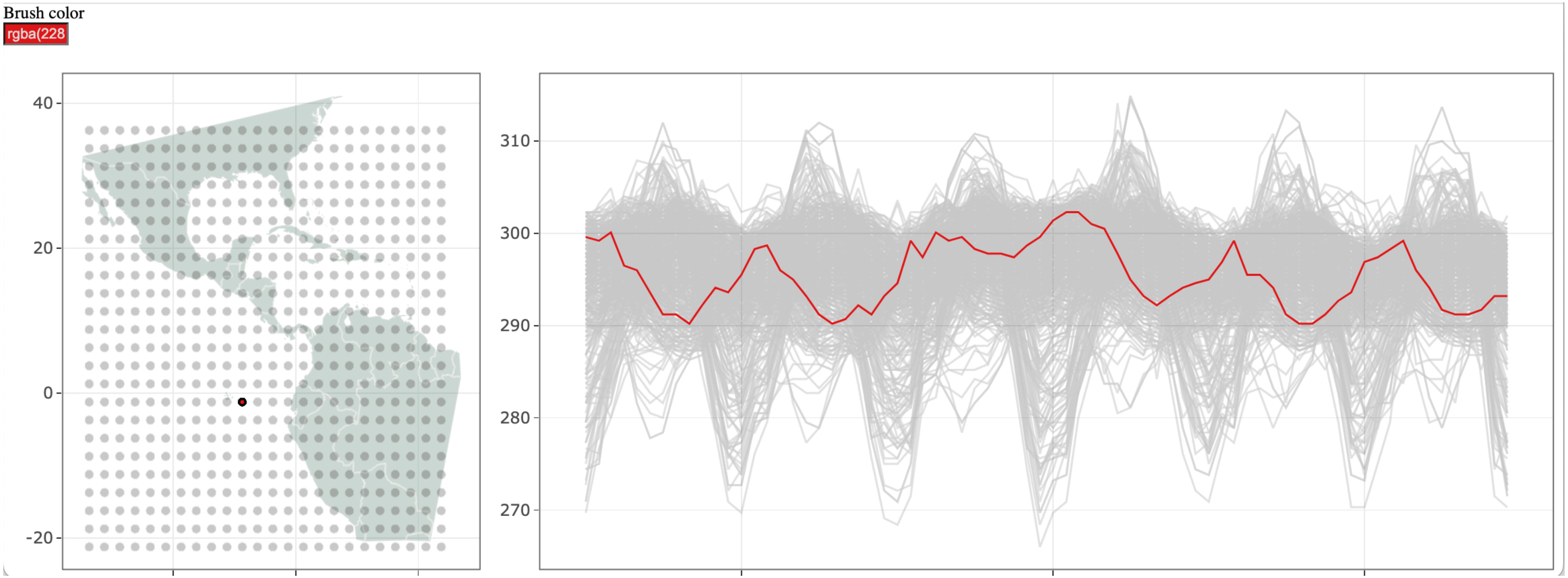


Seasonality can be the focus in the glyphs.

Here monthly min and max temperature over the 6 years is shown, as ribbon glyphs.

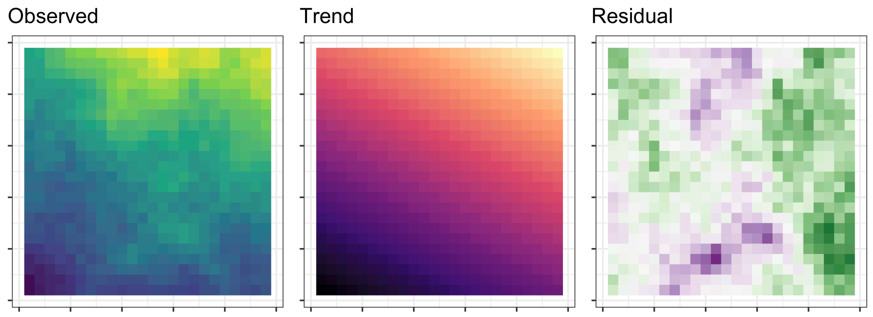
Adding interaction

- DEMO



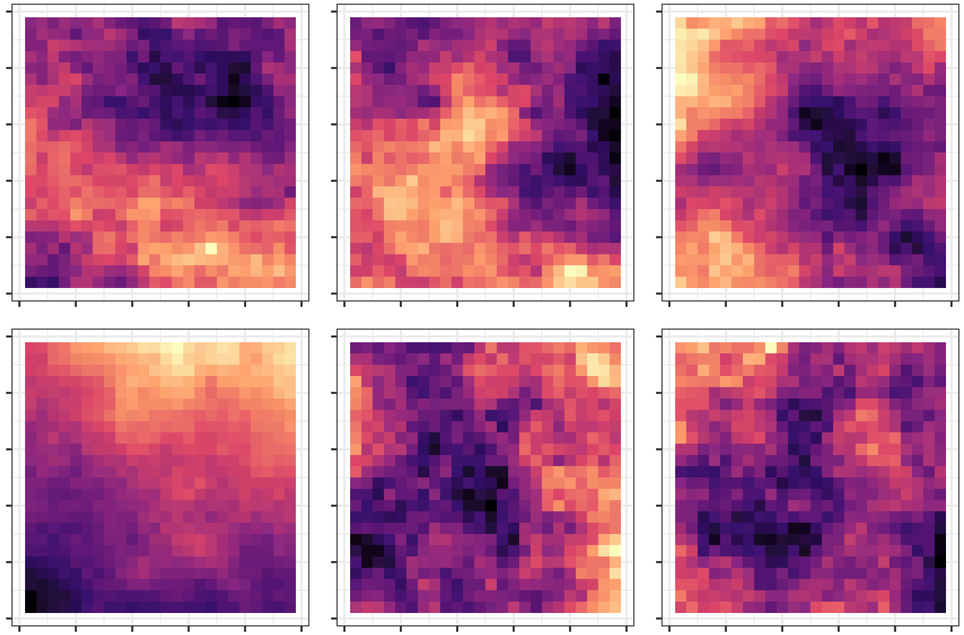
Inference for spatial trend

- generate-data
- plot



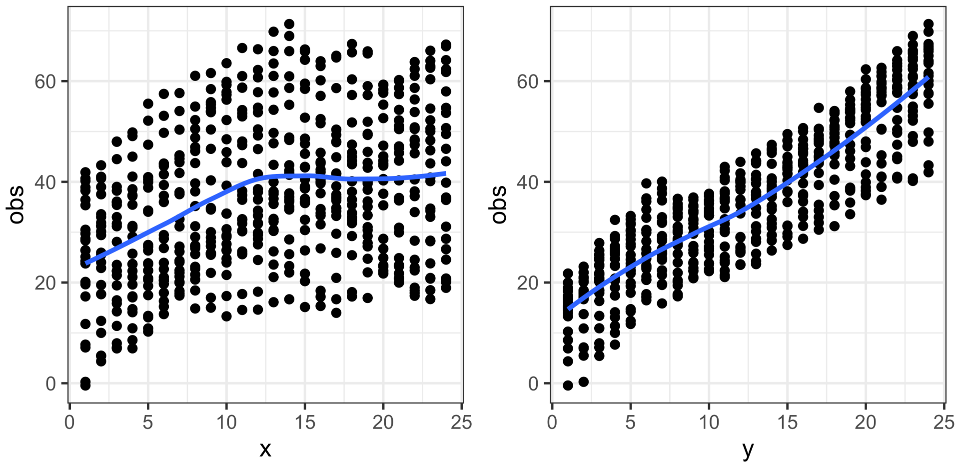
Generate nulls by simulating from the spatial dependence model

- generate-nulls



Extracting spatial trends

- plot



A flash back to the 1970s: Tukey's median polish

This is a useful data scratching technique, particularly for spatial data, to remove complicated trends, as long as they are in spatial marginals.

The median polish is designed for two-way tables. Gridded spatial data is a form of two-way table.

Median polish (1/2)

Export

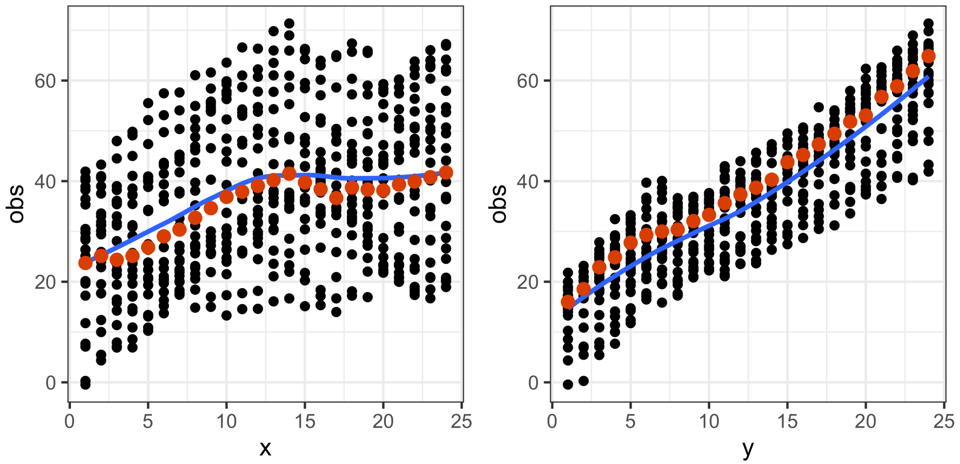
1	2	3
4	5	6
7	8	9

1. Compute overall median and residual table.
2. Compute the row medians.
3. Create a new residual table from the row medians.
4. Compute the column medians.
5. Create a new residual table from the column medians.
6. Second iteration – row effects.
7. Second iteration – column effects
8. Iterate through steps 2-5 until row and column effect medians are close to 0.

Median polish (2/2)

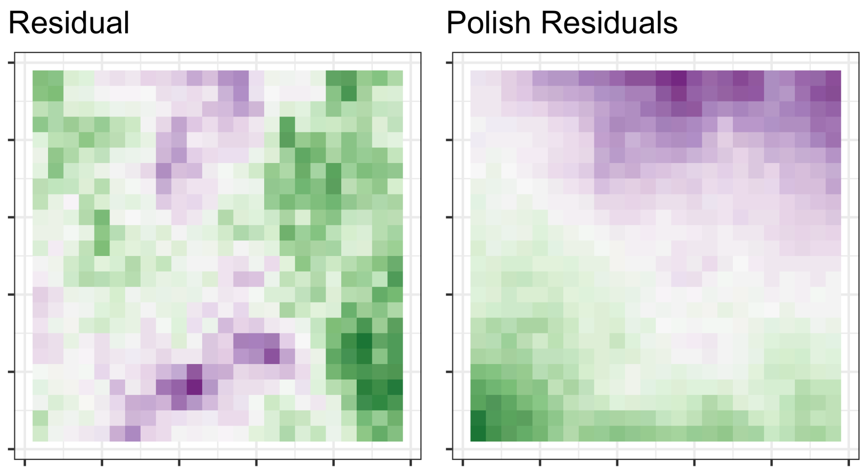
Column and row effects from median polish

▶ Code



Resulting in the residuals as:

▶ Code



Spatial data needs maps

Spatial polygon data

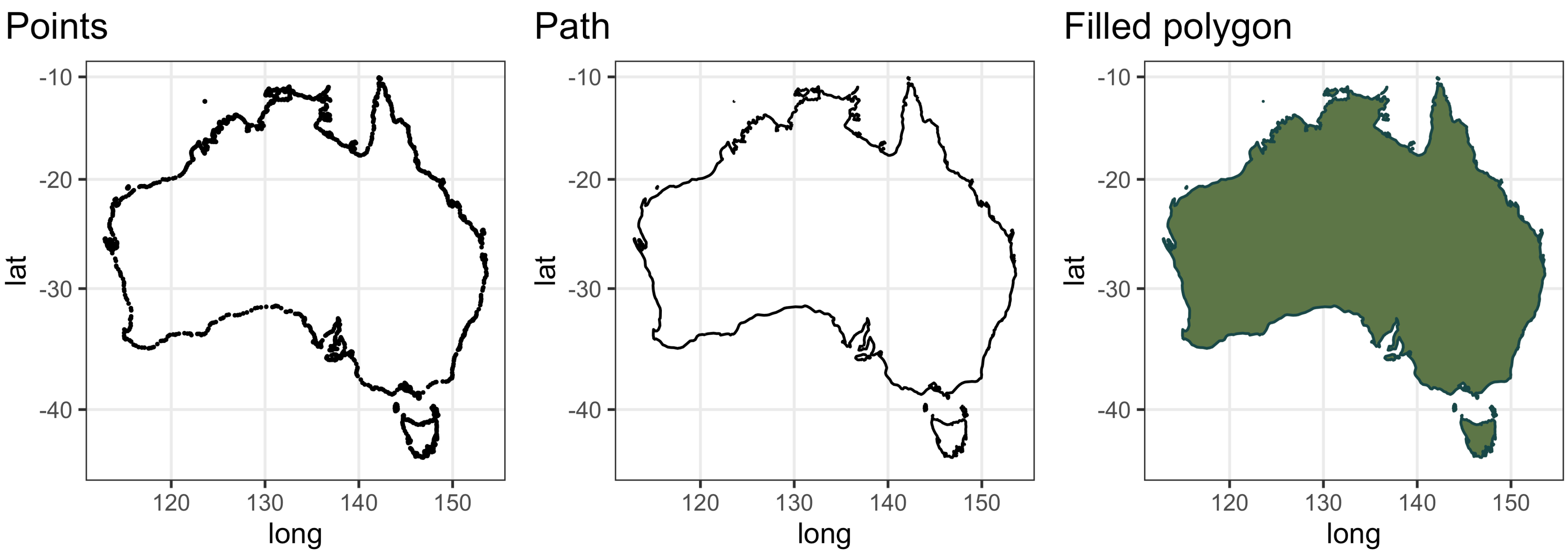
Show 10 entries Search:

	long	lat	group	order	region	subregion
1	123.5945281982422	-12.42568302154541	133	7115	Australia	Ashmore and Cartier Islands
2	123.5952072143555	-12.43593692779541	133	7116	Australia	Ashmore and Cartier Islands
3	123.5731506347656	-12.43418025970459	133	7117	Australia	Ashmore and Cartier Islands
4	123.5724639892578	-12.42392539978027	133	7118	Australia	Ashmore and Cartier Islands
5	123.5945281982422	-12.42568302154541	133	7119	Australia	Ashmore and Cartier Islands
6	158.8787994384766	-54.70976257324219	139	7267	Australia	Macquarie Island
7	158.84521484375	-54.74921798706055	139	7268	Australia	Macquarie Island
8	158.8359375	-54.70400238037109	139	7269	Australia	Macquarie Island
9	158.89697265625	-54.50605392456055	139	7270	Australia	Macquarie Island
10	158.9588775634766	-54.47236251831055	139	7271	Australia	Macquarie Island

Showing 1 to 10 of 2,579 entries Previous 1 2 3 4 5 … 258 Next

Spatial polygon data

- Code



Spatial polygon data, includes measured values (variables) associated with a spatial polygon.

Spatial polygon data

STEP 1: Thin your map!

Most spatial polygon data is large, with high resolution on the polygons.

This makes them SLOW to plot.

For data analysis needs fast plotting, and resolution can be smaller.

`rmapshaper::ms_simplify()` is the best function.

`sf`: Simple spatial polygon objects in R

Has a coordinate system (projection), and bounding box. Supports technically accurate distance calculations between coordinates (on a sphere).

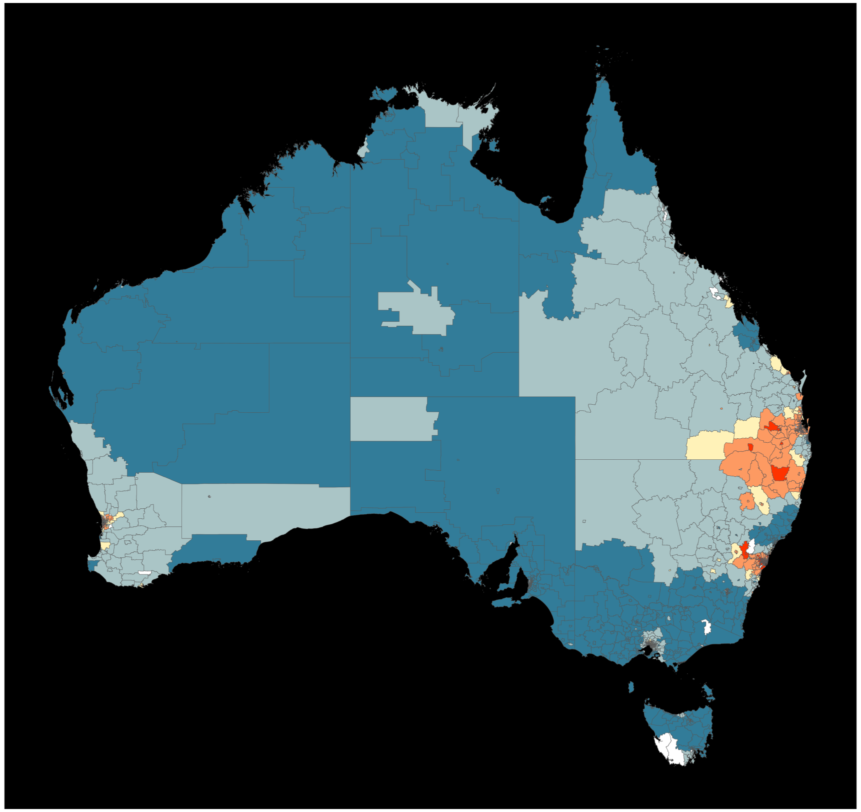
Choropleth maps

A choropleth map is used to show a measured variable associated with a political or geographic region. Polygons for the region are filled with colour.

The purpose is to examine the spatial distribution of a variable.

▶ Code

▶ Code



The problem with choropleth maps

The problem is that high density population areas may be very small geographically. They can disappear in a choropleth map, which means that we get a biased sense of the spatial distribution of a variable.

Cartograms

A cartogram transforms the geographic shape to match the value of a statistic or the population. Its a useful exploratory technique for examining the spatial distribution of a measured variable.

BUT they don't work for Australia.

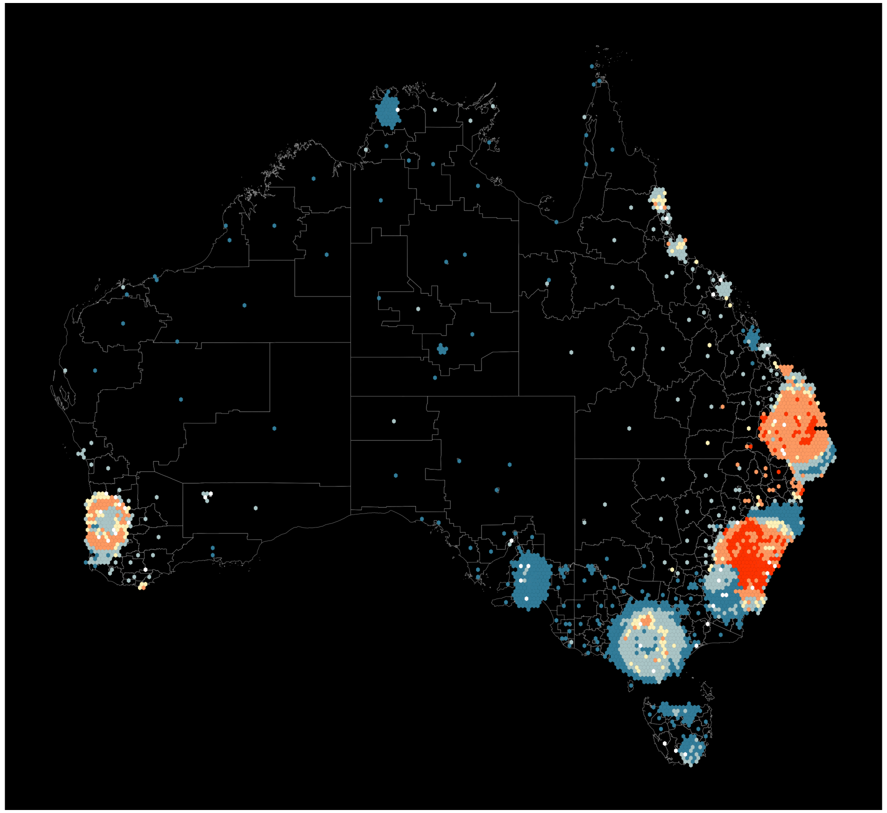
▶ Code

Hexagon tile

A hexagon tile map represents every spatial polygon with an equal sized hexagon. In dense areas these will be tesselated, but separated hexagons are placed at centroids of the remote spatial regions.

It’s not perfect, but now the higher incidence in Perth suburbs, some melbourne suburbs, and Sydney are more visible.

▶ Code



Resources

- cubble: A Vector Spatio-Temporal Data Structure for Data Analysis
- sf: Simple Features for R
- Healy (2018) Data Visualization
- Visualising spatial data using R
- Kobakian et al Hexagon tile map
- Wikle, Zammit-Mangion, Cressie (2018) Spatio-Temporal Statistics with R
- Moraga, Paula. (2019). Geospatial Health Data