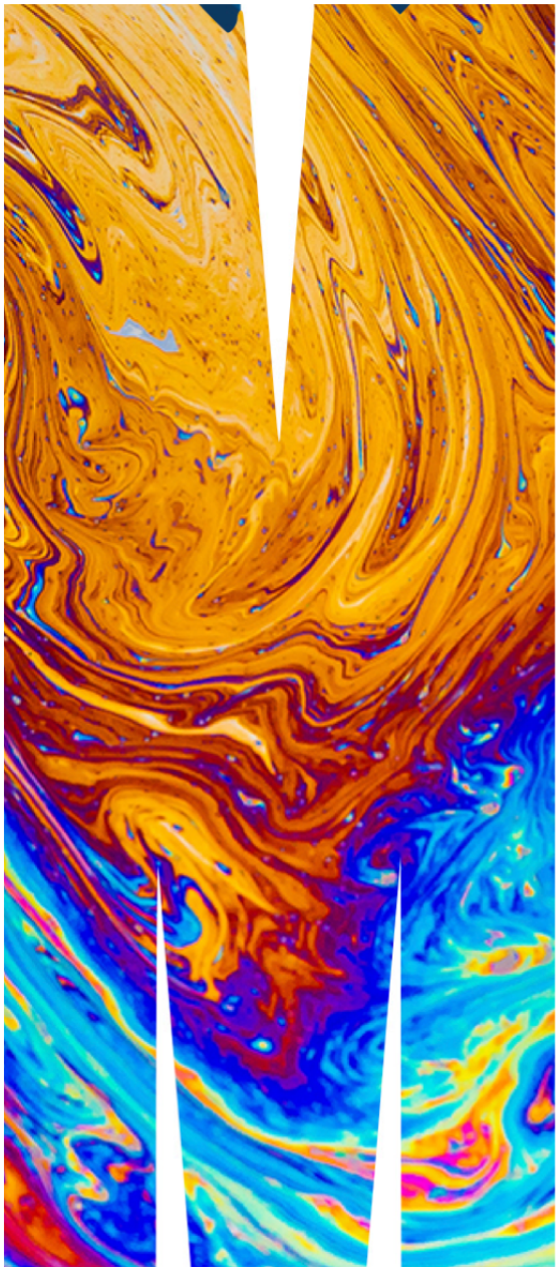
MONASH University

ETC5521: Exploratory Data Analysis

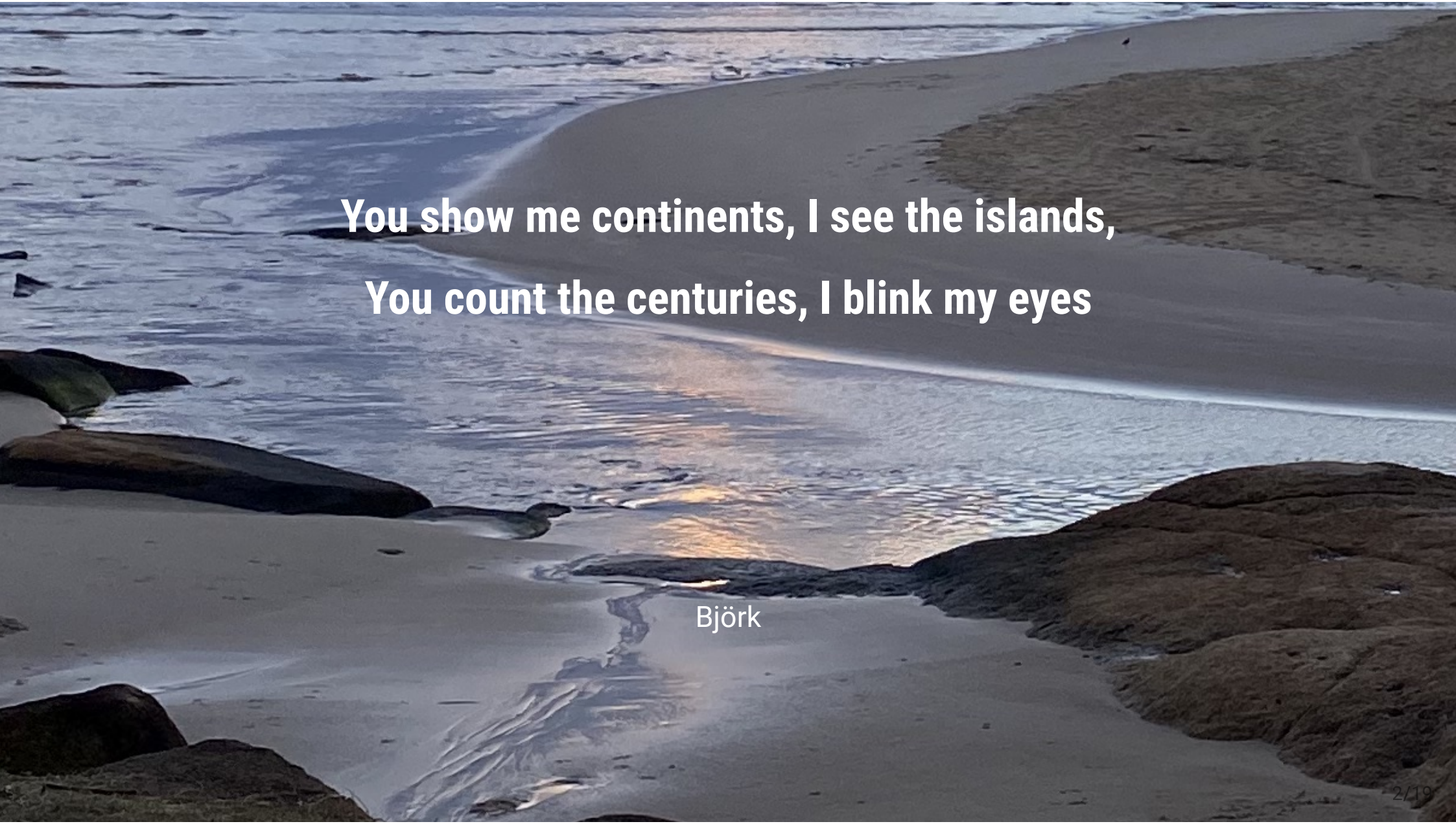
Exploring data having a space and time context



Lecturer: *Di Cook*

ETC5521.Clayton-x@monash.edu

Week 10 - Session 1


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

Outline

First part

- Breaking up data by time, and by space
- Maps of space over time
- Exploring time over space with glyph maps

Second part

- Capturing spatial trend
- Bending the choropleth map
- A flash back to the 1970s: Tukey's median polish

Case study 1 Temperature change in Americas

`data` R

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

time	y	x	lat	long	date	cloudhigh	cloudlow	cloudmid	ozone	pressure	surftemp	temperature	id	day	month	year
1	1	1	-21.2	-113.80000	1995-01-01	0.5	31.0	2.0	260	1000	297.4	296.9	1-1	0	1	1995
1	1	2	-21.2	-111.29565	1995-01-01	1.5	31.5	2.5	260	1000	297.4	296.5	2-1	0	1	1995
1	1	3	-21.2	-108.79130	1995-01-01	1.5	32.5	3.5	260	1000	297.4	296.0	3-1	0	1	1995
1	1	4	-21.2	-106.28696	1995-01-01	1.0	39.0	4.0	258	1000	296.9	296.5	4-1	0	1	1995
1	1	5	-21.2	-103.78261	1995-01-01	0.5	48.0	4.5	258	1000	296.5	295.5	5-1	0	1	1995
1	1	6	-21.2	-101.27826	1995-01-01	0.0	50.0	2.5	258	1000	296.5	295.0	6-1	0	1	1995
1	1	7	-21.2	-98.77391	1995-01-01	0.0	51.0	4.5	256	1000	295.5	295.5	7-1	0	1	1995
1	1	8	-21.2	-96.26957	1995-01-01	0.0	52.5	5.0	258	1000	295.5	295.0	8-1	0	1	1995
1	1	9	-21.2	-93.76522	1995-01-01	0.5	54.0	8.5	256	1000	295.0	295.0	9-1	0	1	1995
1	1	10	-21.2	-91.26087	1995-01-01	1.0	56.0	11.5	258	1000	294.6	294.6	10-1	0	1	1995

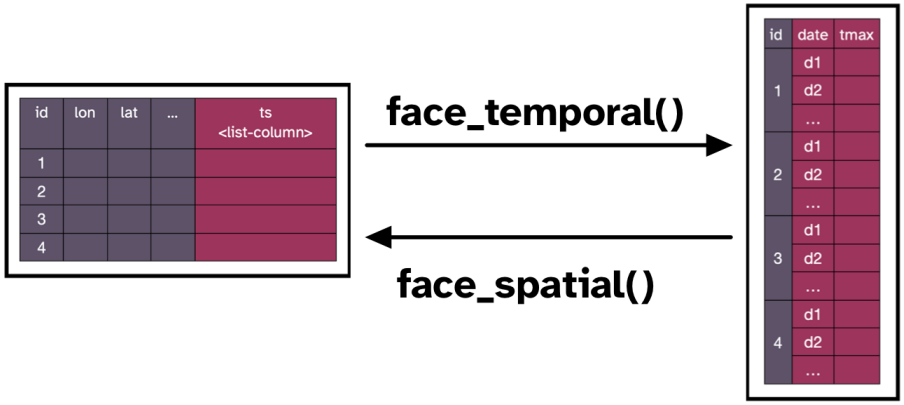
Spatiotemporal object in R: cubble

```
nasa_cb <- as_cubble(as_tibble(nasa),
                     key=id,
                     index=time,
                     coords=c(long, lat))
nasa_cb
```

```
## # cubble:   key: id [576], index: time, nested form
## # spatial:  [-113.8, -21.2, -56.2, 36.2], Missing (
## # temporal: time [int], date [dttm], cloudhigh [dbl
##        y     x   lat   long id    ts
##    <int> <int> <dbl>  <dbl> <chr> <list>
##  1     1     1 -21.2 -114.  1-1   <tibble [72 × 12]
##  2     1     2 -21.2 -111.  2-1   <tibble [72 × 12]
```

```
nasa_cb %>% face_temporal()
```

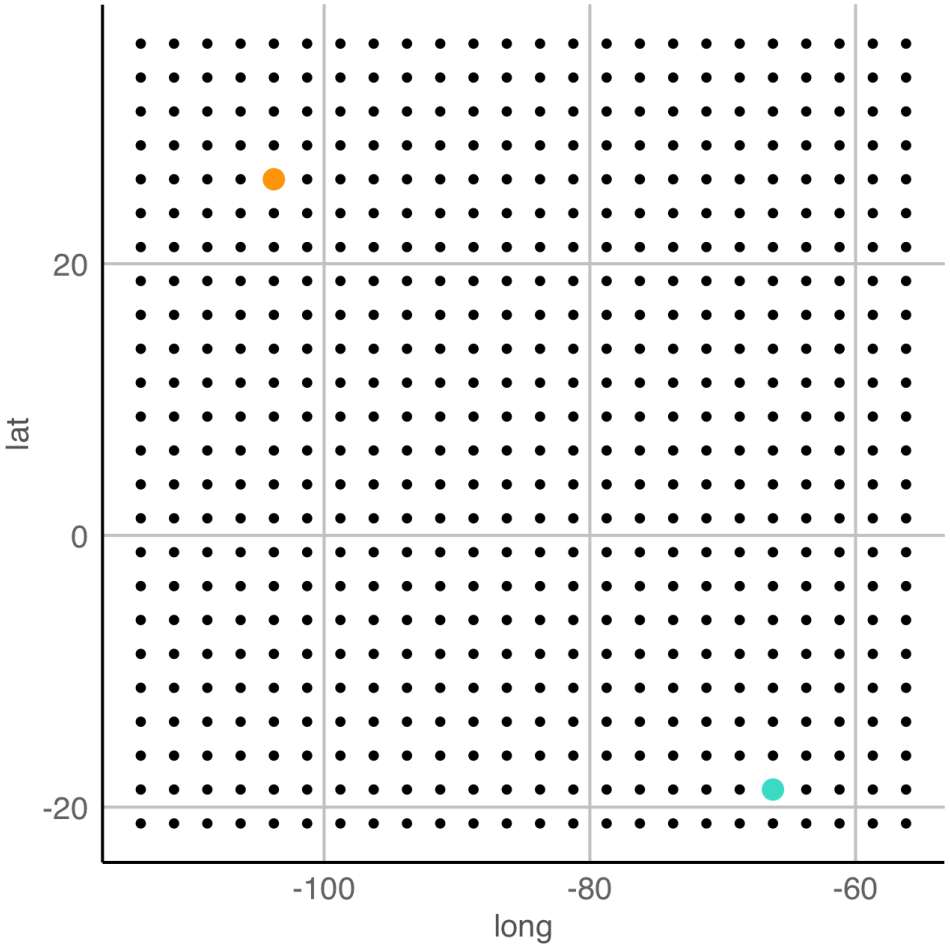
```
## # cubble:   key: id [576], index: time, long form
## # temporal: 1 -- 72 [1], no gaps
## # spatial:  y [int], x [int], lat [dbl], long [dbl]
##    id     time date                cloudhigh cloudl
##    <chr> <int> <dttm>                  <dbl>    <db
##  1 1-1       1 1995-01-01 00:00:00       0.5     31
##  2 1-1       2 1995-02-01 00:00:00       1       33
##  3 1-1       3 1995-03-01 00:00:00       2       25
##  4 1-1       4 1995-04-01 00:00:00       4       33
##  5 1-1       5 1995-05-01 00:00:00       6.5     36
##  6 1-1       6 1995-06-01 00:00:00       3.5     36
```



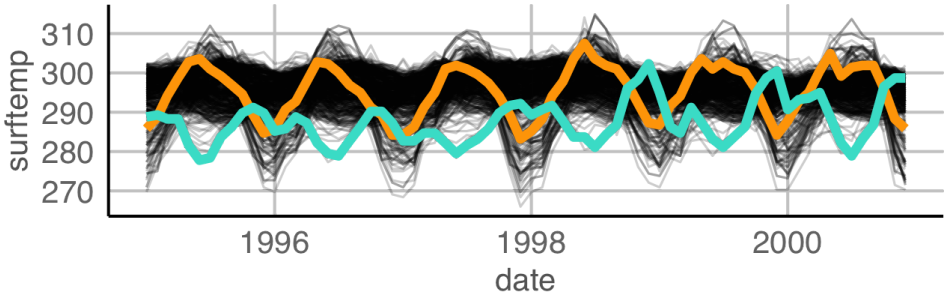
Like a `tibble` but can pivot back and forth between spatial and temporal components.

Case study 1 Temperature change in Americas

plot R



plot R



Pre-processing of time and space

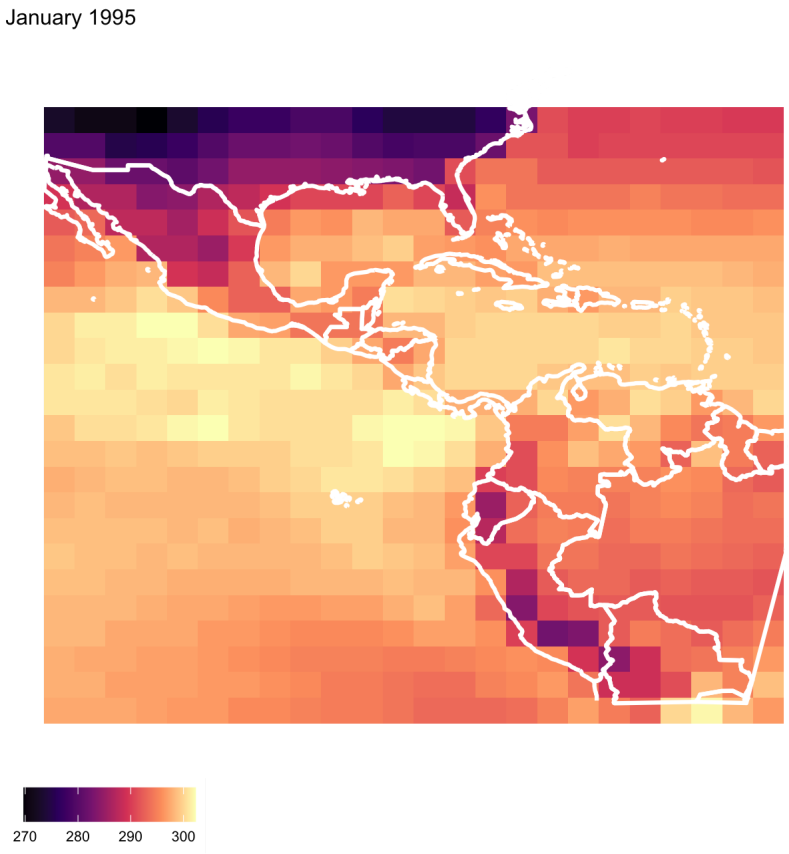
Think of time and space as ordered categorical variables.

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

For the nasa data, this is already done. 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.

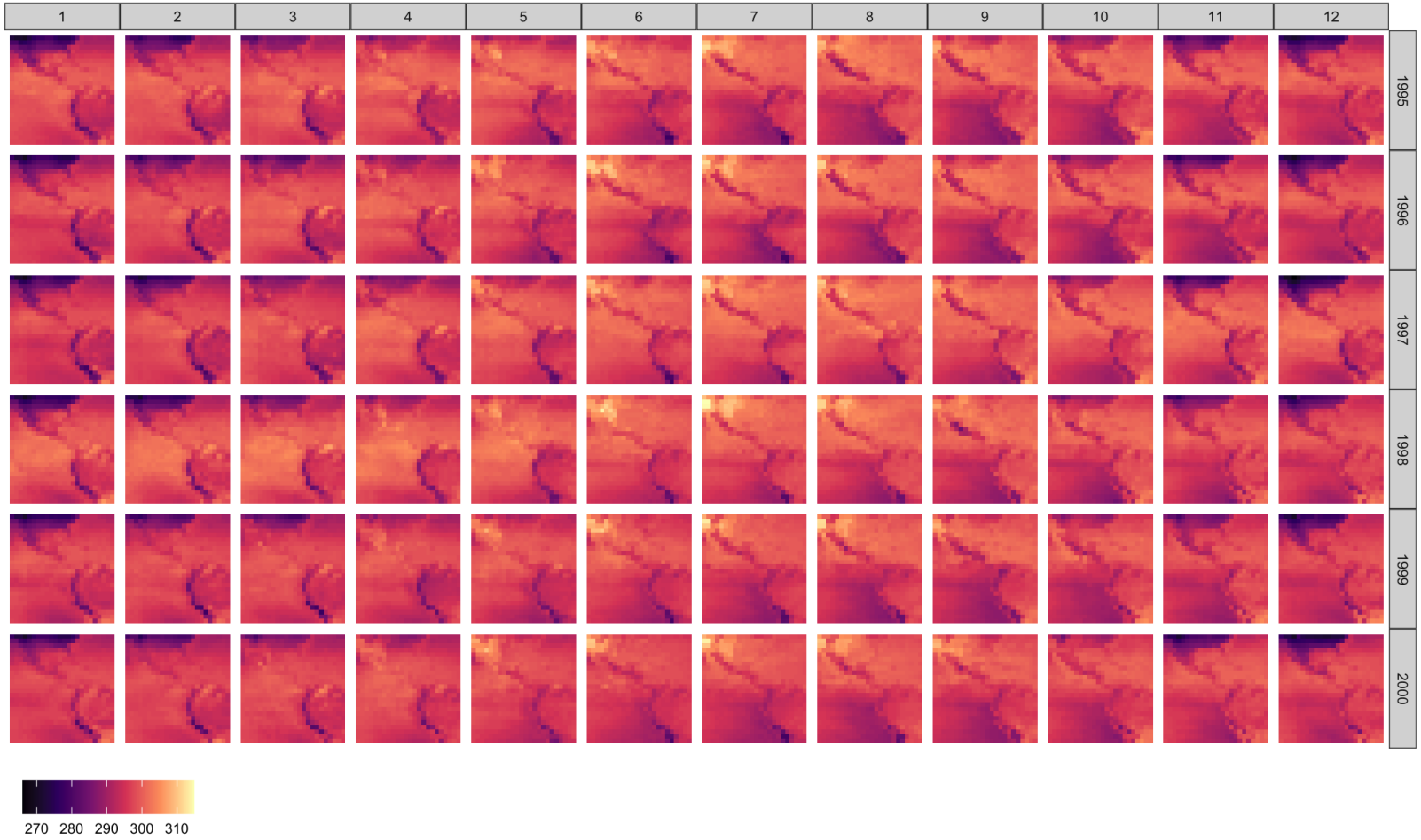
Slice in time and create a spatial map

```
plot    learn    R
```



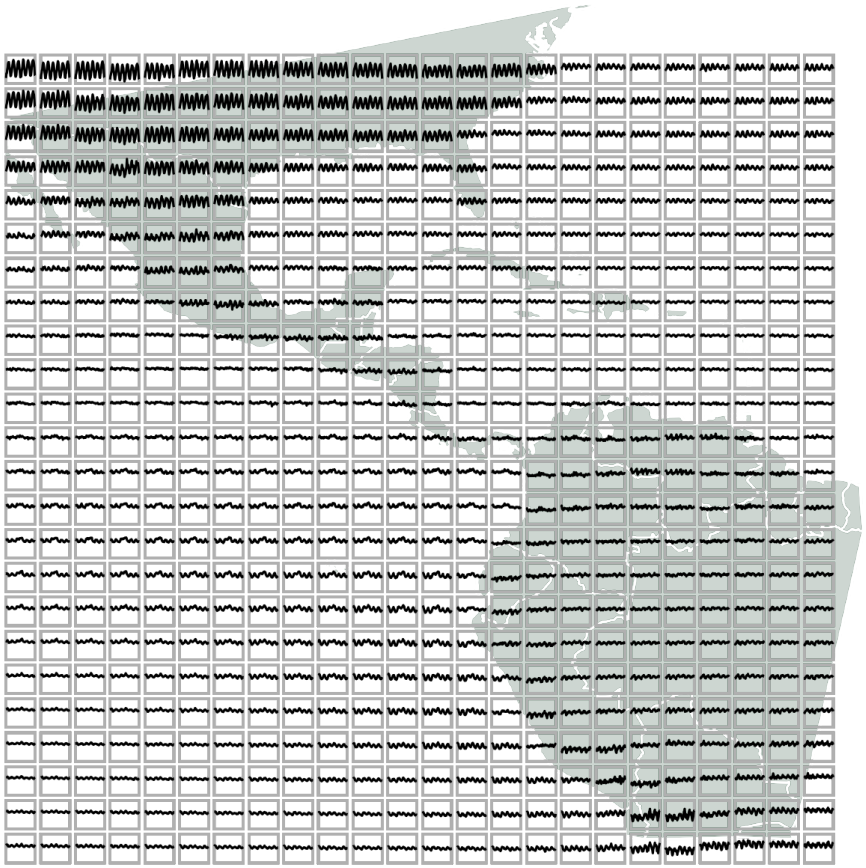
Explore spatial trend over time

`plot` `learn` `R`

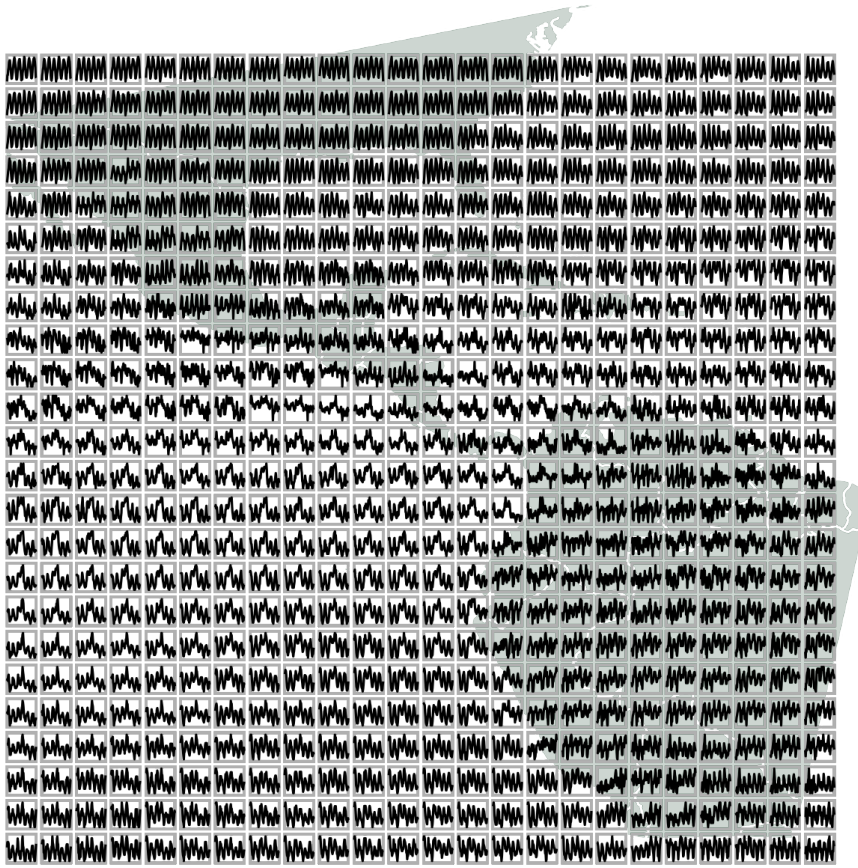


glyphmap: time across space

plot learn R

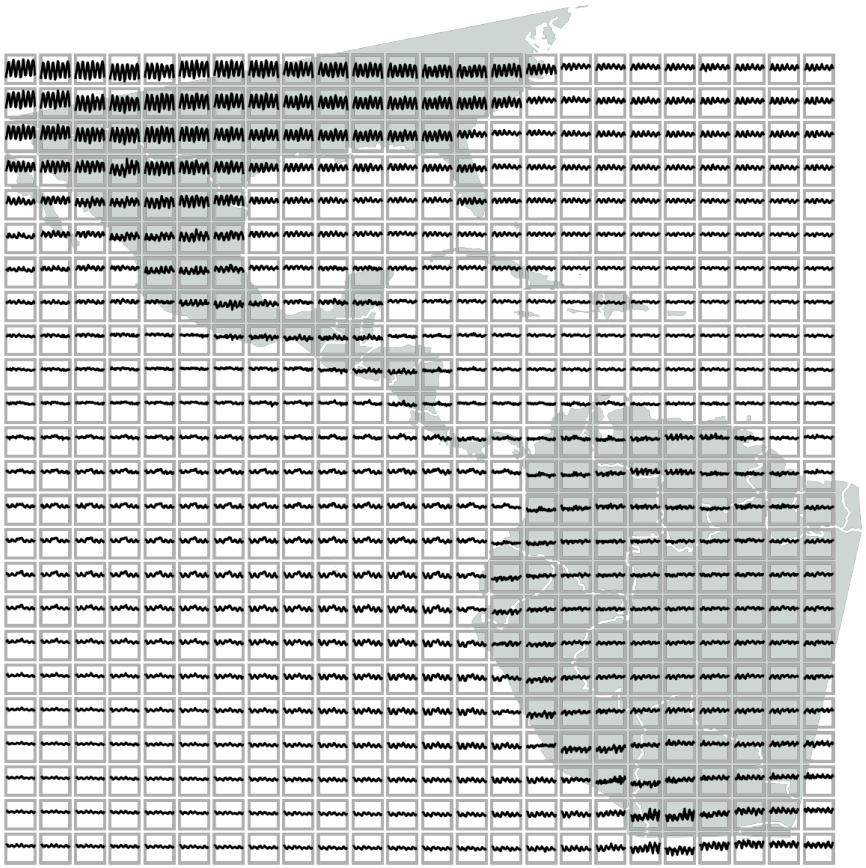


plot learn R

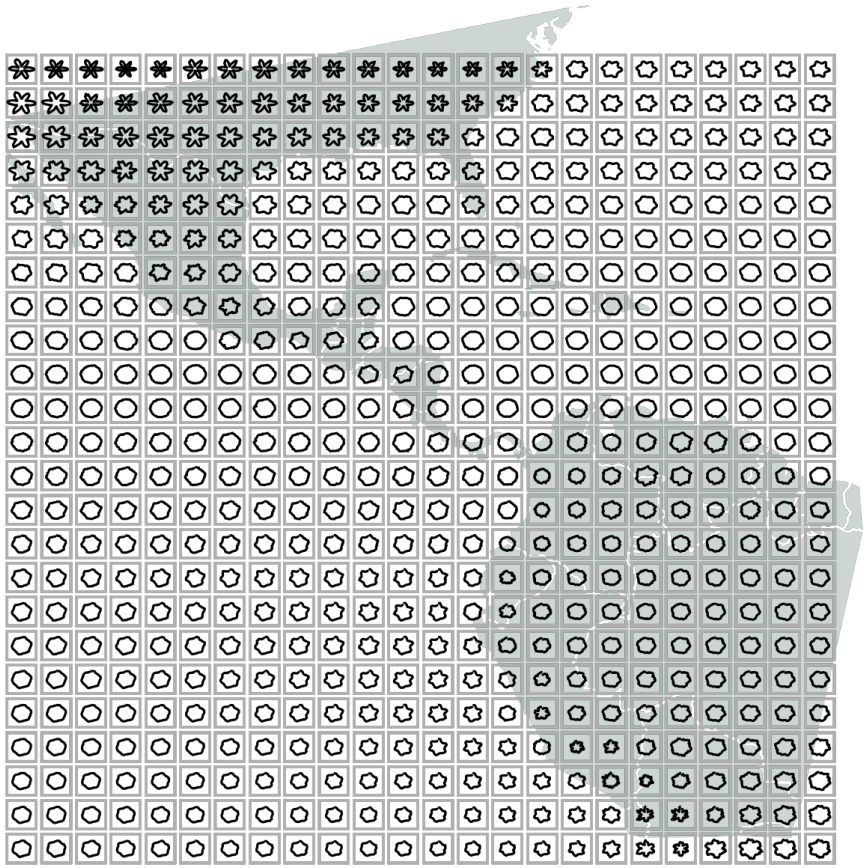


glyphmap: time across space

plot learn R



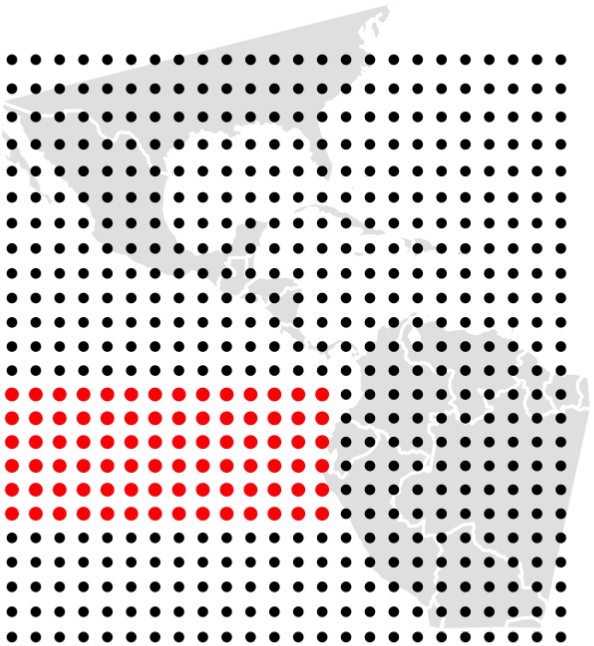
plot learn R

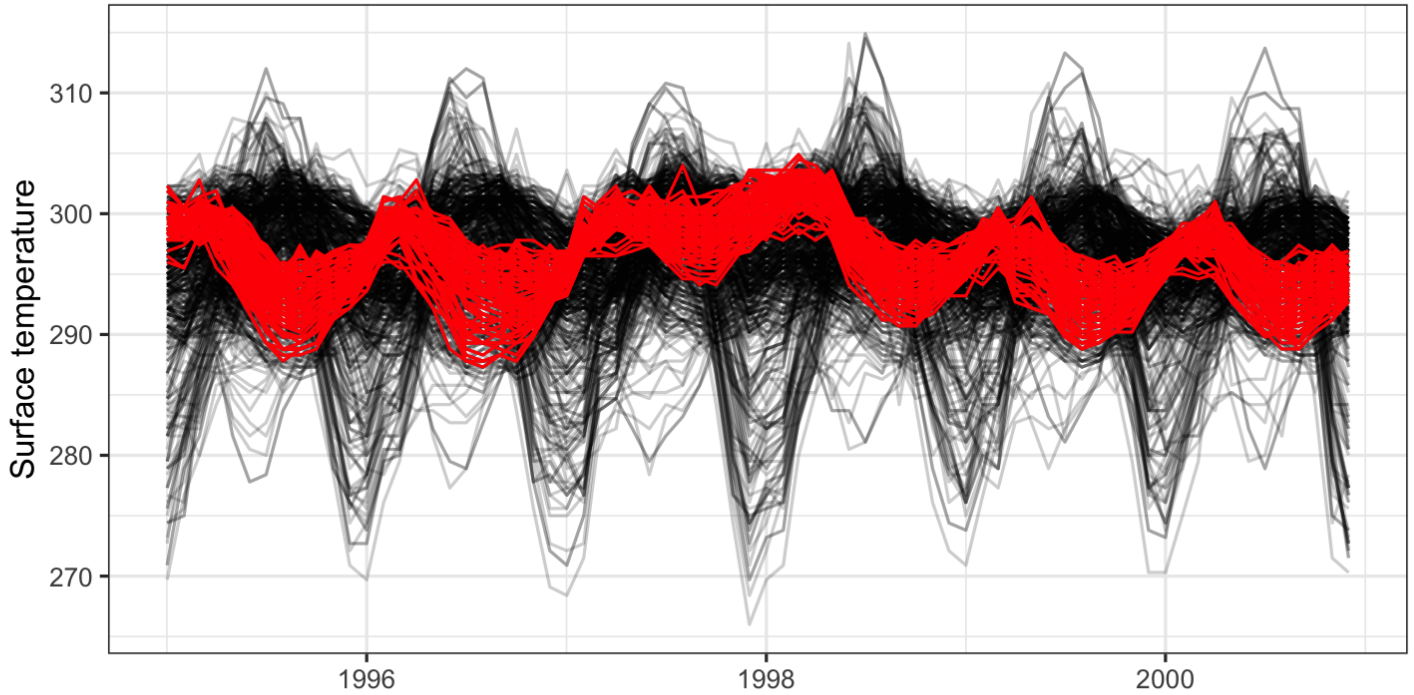


Case study 1 Temperature change in Americas

Exploring El Nino

Slice space, and show the time series, and the pattern is very clear: The seasonal water temperature decrease doesn't happen in 1997, and water in this area stays unseasonably warm.





🔧 Your turn using tsibbletalk

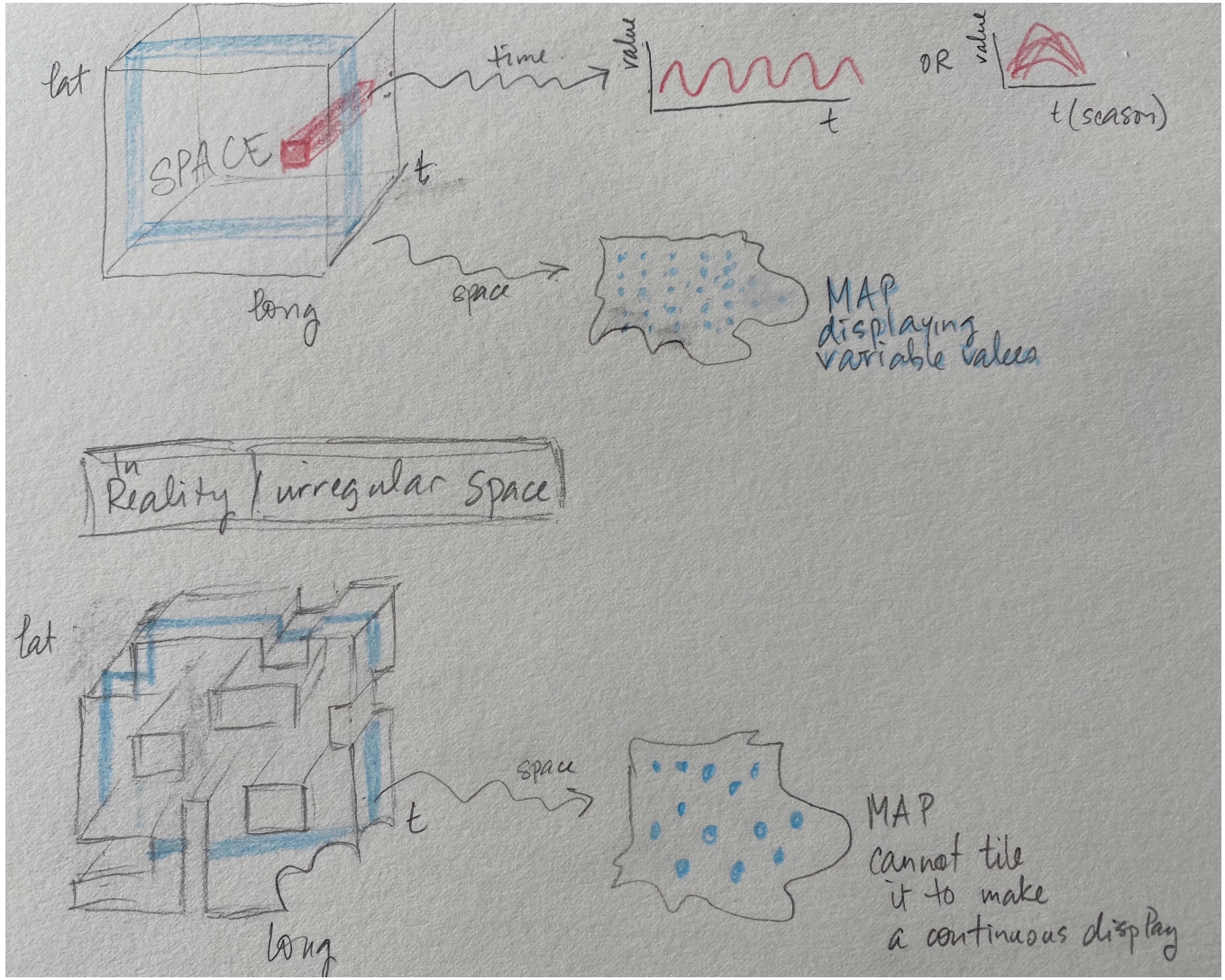
```
library(tsibble)
library(tsibbletalk)
library(lubridate)
nasa_shared <- nasa %>%
  mutate(date = ymd(date)) %>%
  select(long, lat, date, surftemp, id) %>%
  as_tsibble(index=date, key=id) %>%
  as_shared_tsibble()
p1 <- nasa_shared %>%
  ggplot(aes(x = long, y = lat)) +
  geom_point(aes(group = id))
p2 <- nasa_shared %>%
  ggplot(aes(x = date, y = surftemp)) +
  geom_line(aes(group = id), alpha = 0.5)
library(plotly)
subplot(
    ggplotly(p1, tooltip = "Region", width = 100),
    ggplotly(p2, tooltip = "Region", width = 900),
    nrows = 1, widths=c(0.4, 0.6)) %>%
  highlight(dynamic = TRUE)
```

Thinking about spatiotemporal data

Space is a continuous variable, and in theory it fills out a square (or polygon).

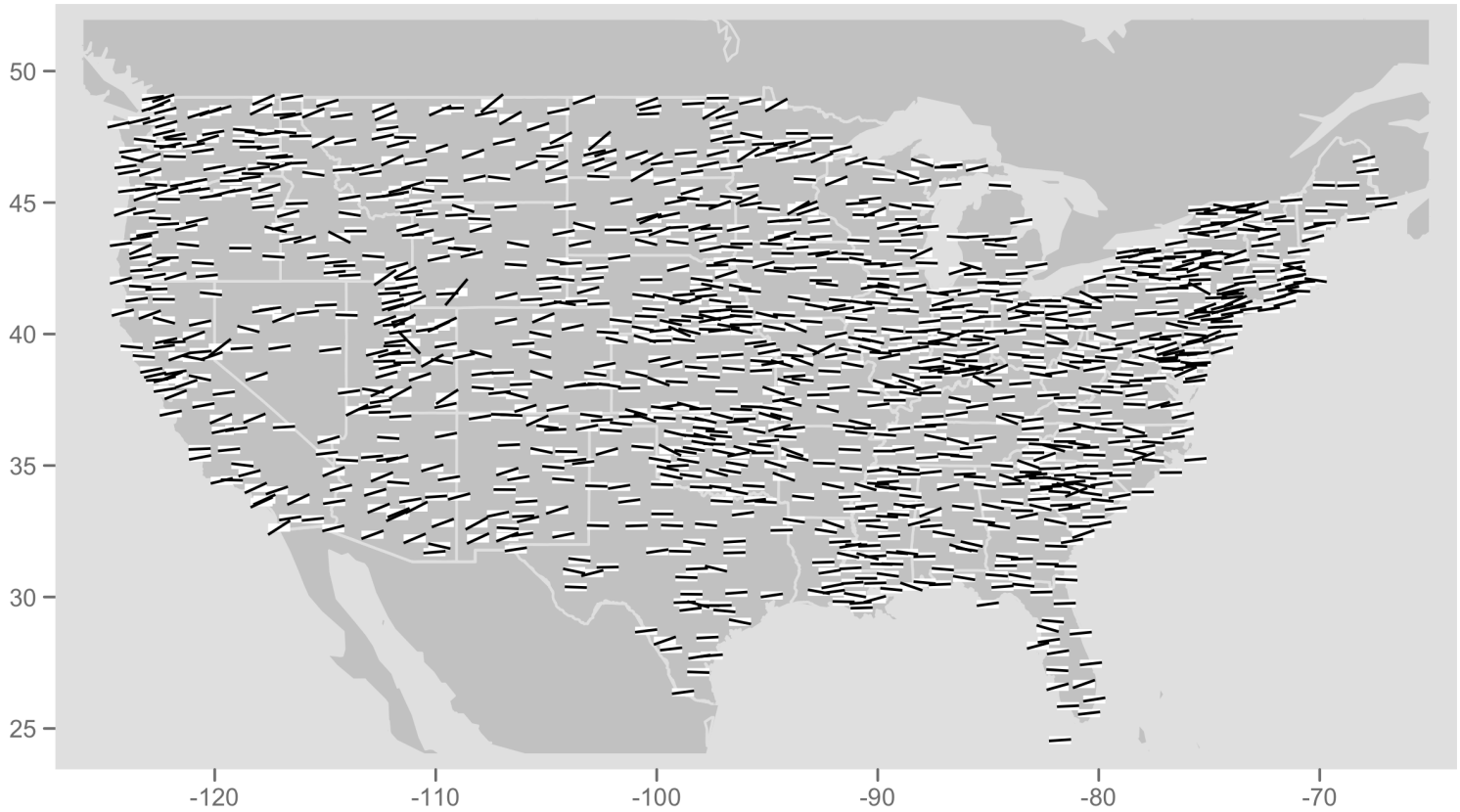
Often, though it is measured irregularly. Sensor stations are sporadically placed.

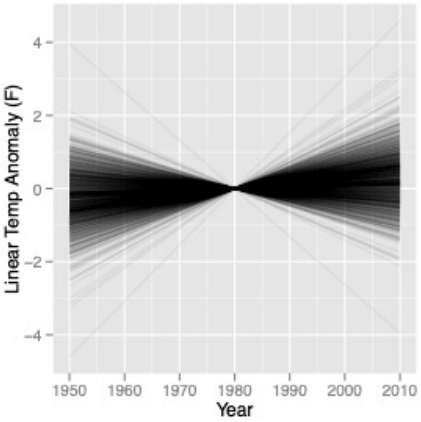
This affects making a density display of the measured variable. Strategy is to plot it at the resolution given, before trying to make a regular grid.



Glyphmaps of irregular space data

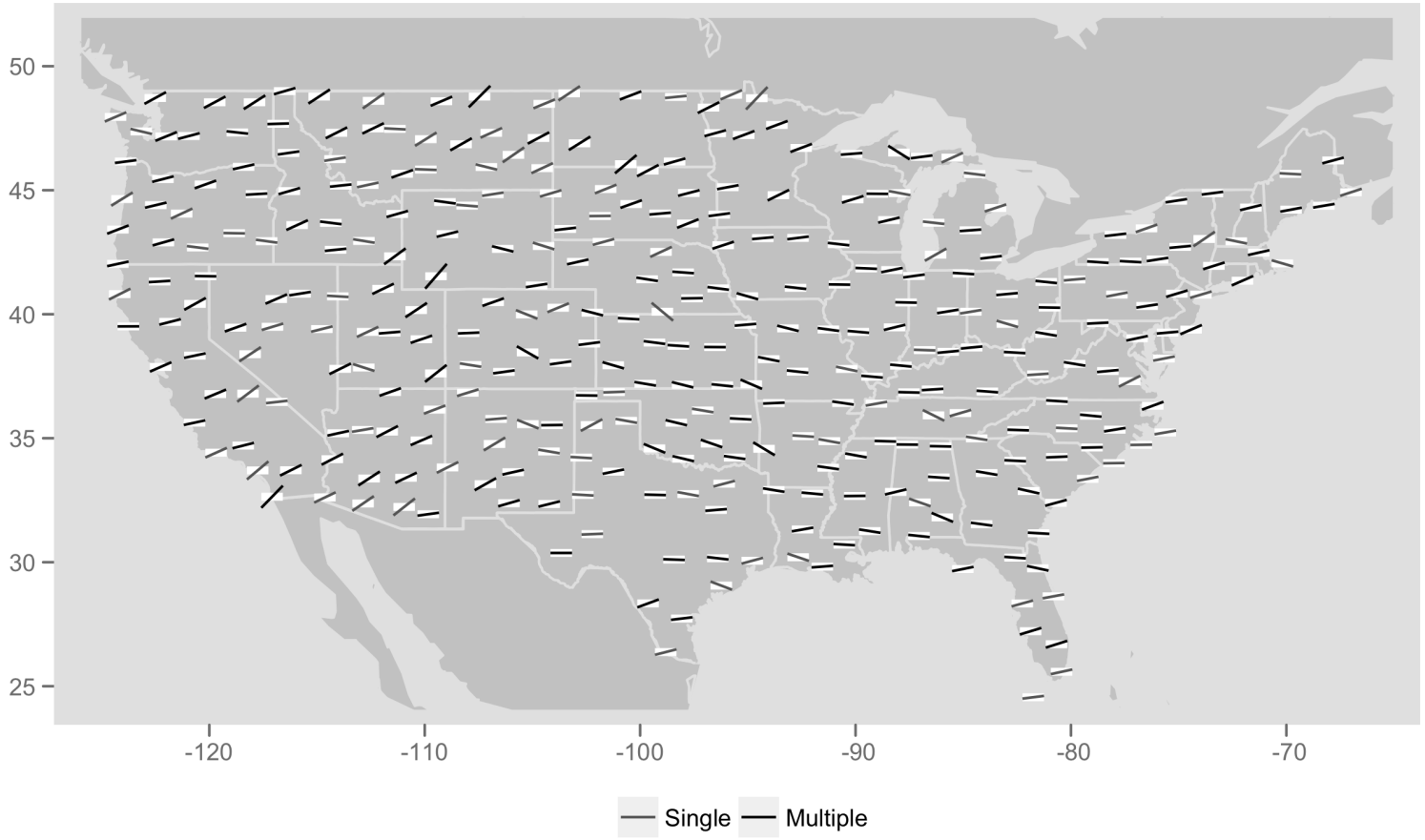
Linear model fit to temperature recorded at historical weather stations across the USA (USHCN)

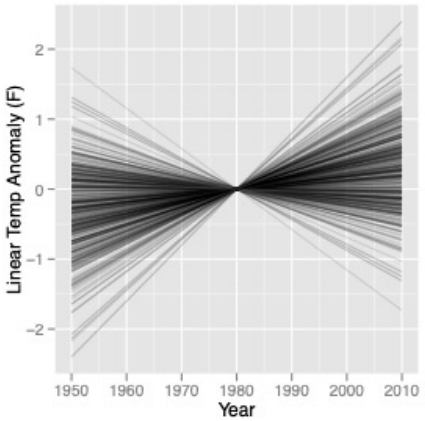




Regularising

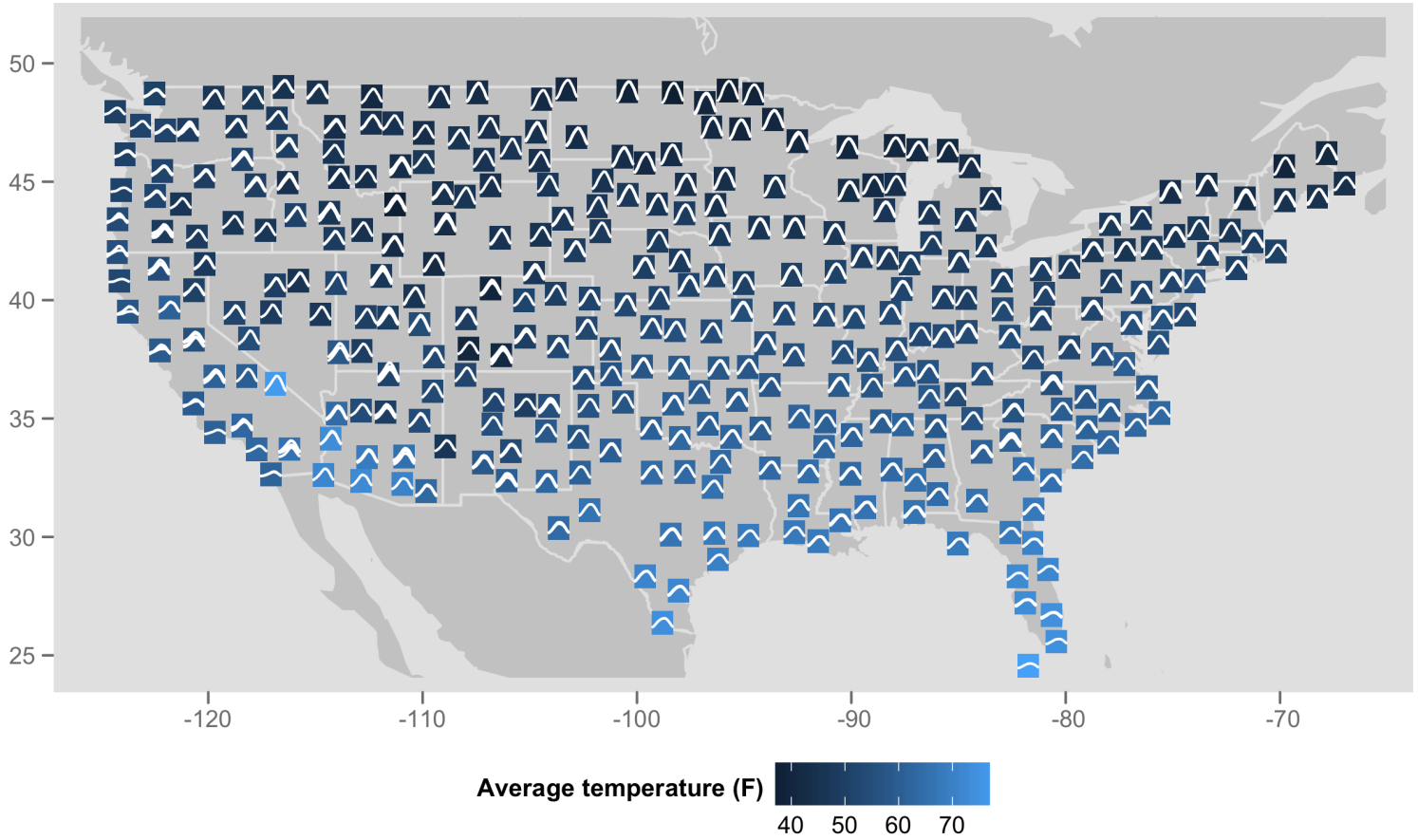
Measurements from nearby stations have been merged. Note, some areas of the USA have been cooling. This is interesting!

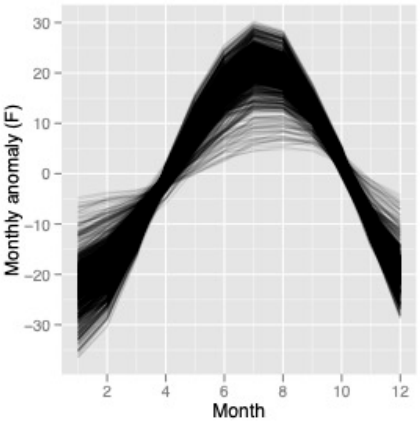




Seasonality

De-construct time into a seasonal component, and representing this on the map is reasonable also.

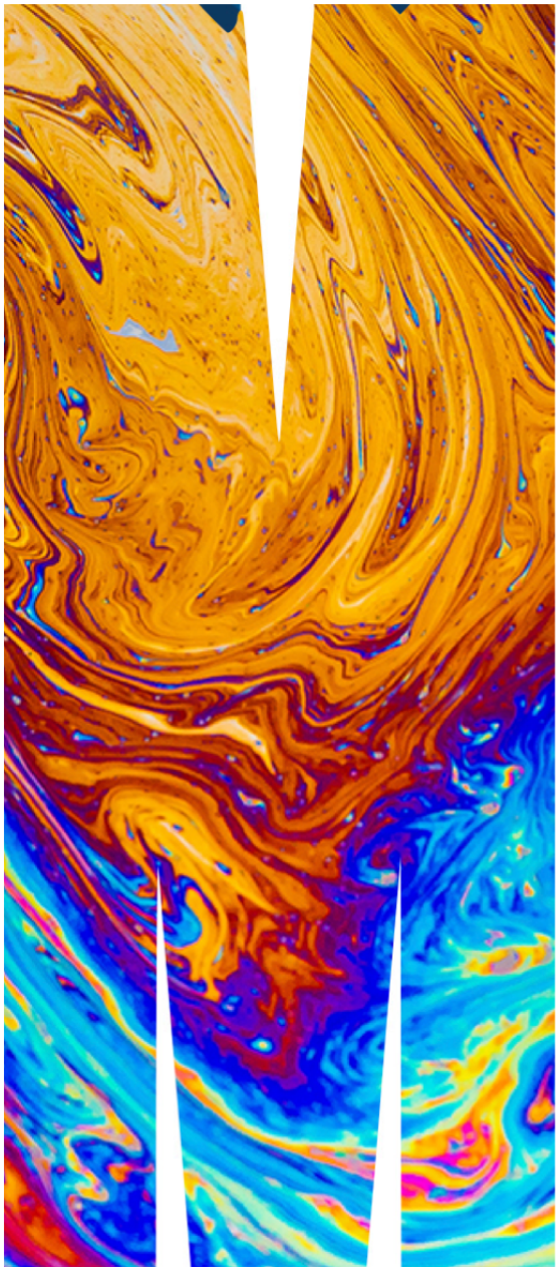




Resources and Acknowledgement

- cubble: A Vector Spatio-Temporal Data Structure for Data Analysis
- sf: Simple Features for R
- Healy (2018) Data Visualization
- Perpinan Lamigueiro (2018) Displaying time series, spatial and space-time data with R
- Wikle, Zammit-Mangion, Cressie (2018) Spatio-Temporal Statistics with R
- Moraga, Paula. (2019). Geospatial Health Data
- Visualising spatial data using R

MONASH University



Lecturer: *Di Cook*

ETC5521.Clayton-x@monash.edu

Week 10 - Session 1