WOMBAT 2025 Tutorial

Visualising Uncertainty

Harriet Mason, Dianne Cook
Department of Econometrics and Business Statistics

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Session 2
Diving deeper into uncertainty visualisation using examples in spatial data

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Introduction to Spatial Visualisation

Why focus on spatial visualisations?

- Spatial case is a good example to work through because the aesthetics we have to express estimates are limited
- Maps take up most of the usual aesthetics by being a representation of space
  - position, size, shape, etc. all have an implicit meaning in the mapping context
  - colour/fill is usually the only aesthetic we have left
  - can also get creative and do glyph maps (we will ignore this variation here)

> ⓘ Spatial data takes up the two dimensions of the display leaving colour and fill to map uncertainty.

Example: Citizen Scientist Data

- There have been reports of a strange spatial pattern in the temperatures of Iowa
- We get some citizen scientists to measure data at their home and report back
- To maintain anonymity, we are only provided with the county of each scientist

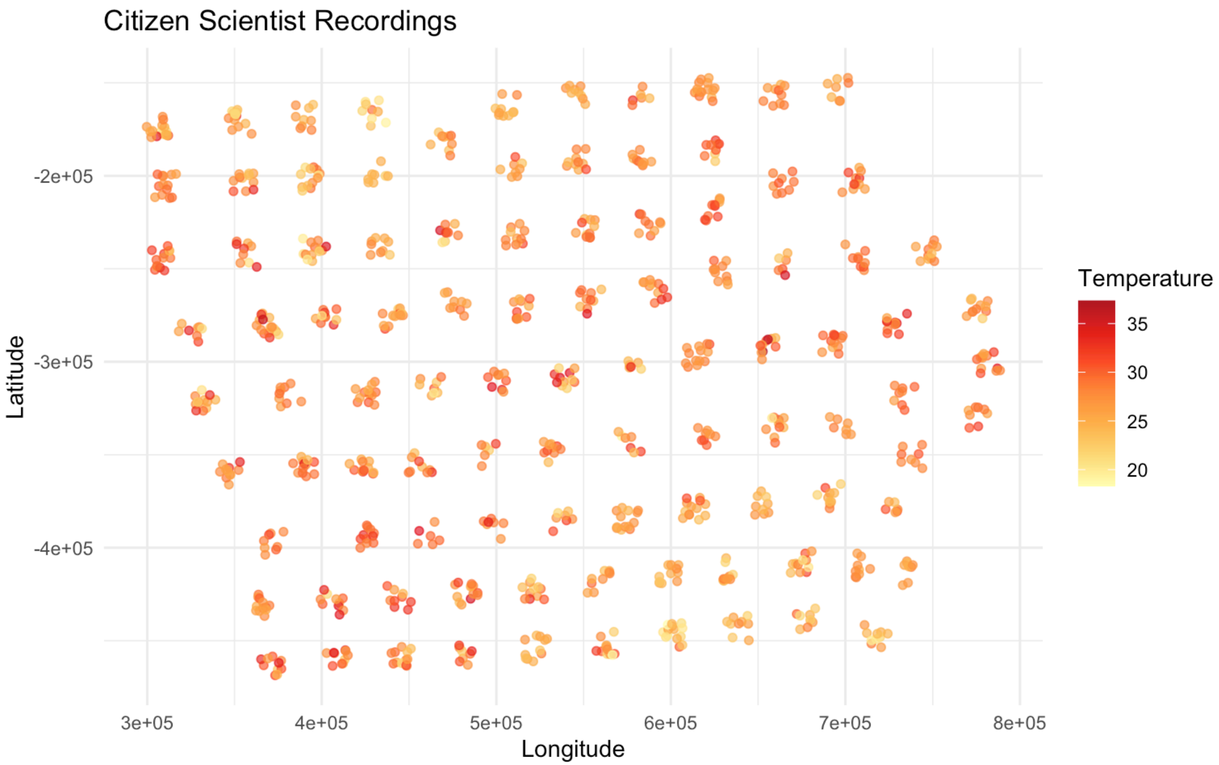
scientistID	county_name	recorded_temp
#74991	Lyon County	21.1
#22780	Dubuque County	28.9
#55325	Crawford County	26.4
#46379	Allamakee County	27.1
#84259	Jones County	34.2

`990` citizen scientists participated

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

We could just plot the data…

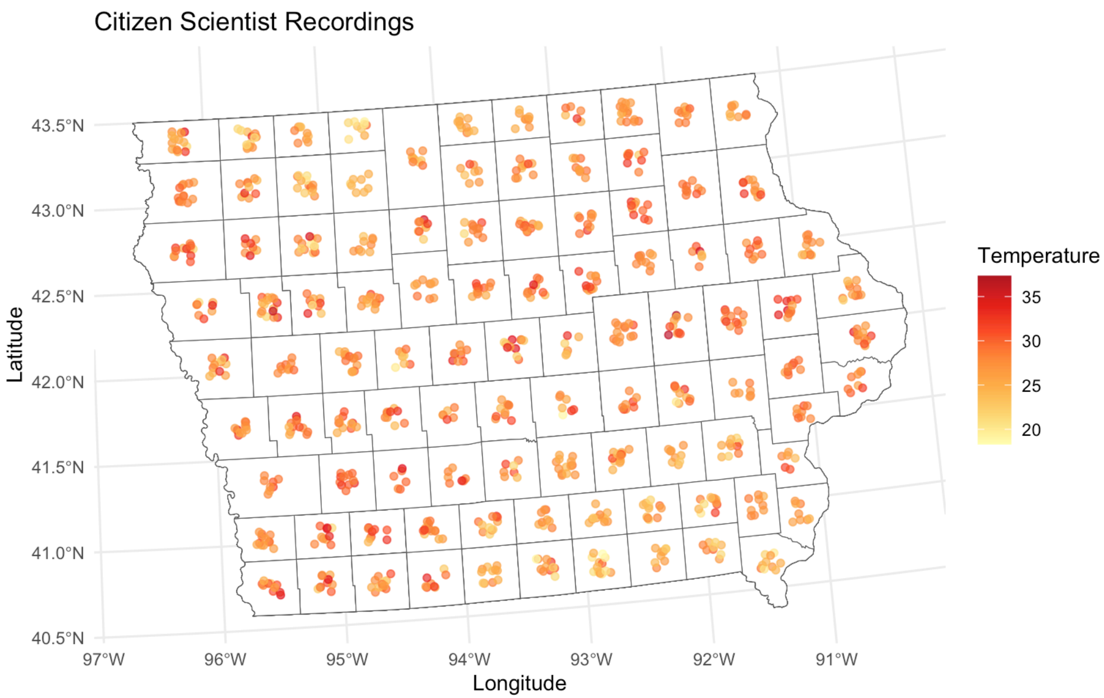
- We often can plot the longitude and latitude directly using `geom_point`.
- While this approach has a low barrier to entry, it lacks the contextual information that gives our plots meaning.



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

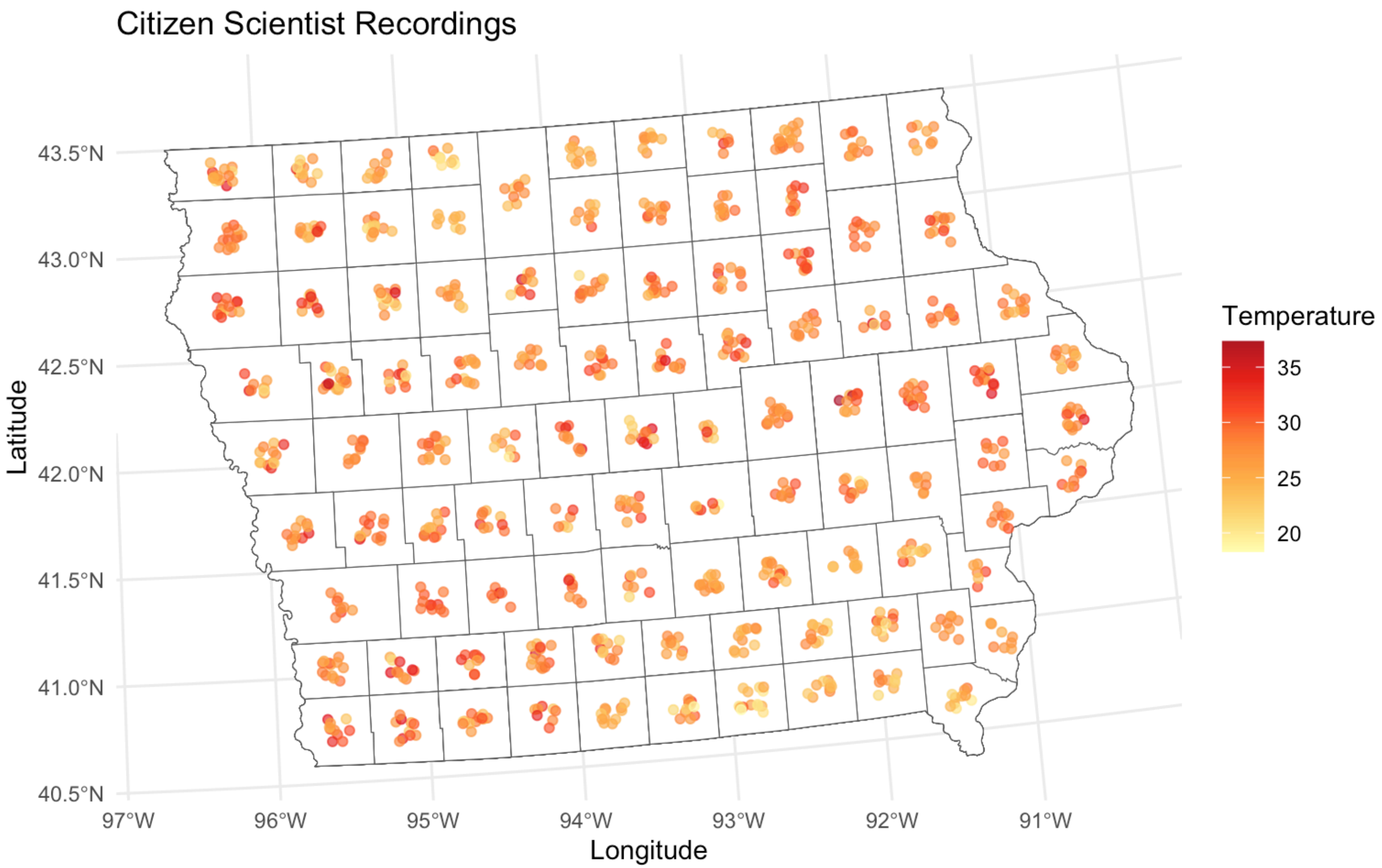
Spatial features objects

- SF objects are differentiated from our usual `tibble` by the additional metata in the Coordinate reference system (CRS)
  - Assumptions about the shape of the planet (geodetic datum)
  - Distortions we will/won't accept when drawing the map (map projection)



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Can you see the spatial trend?



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

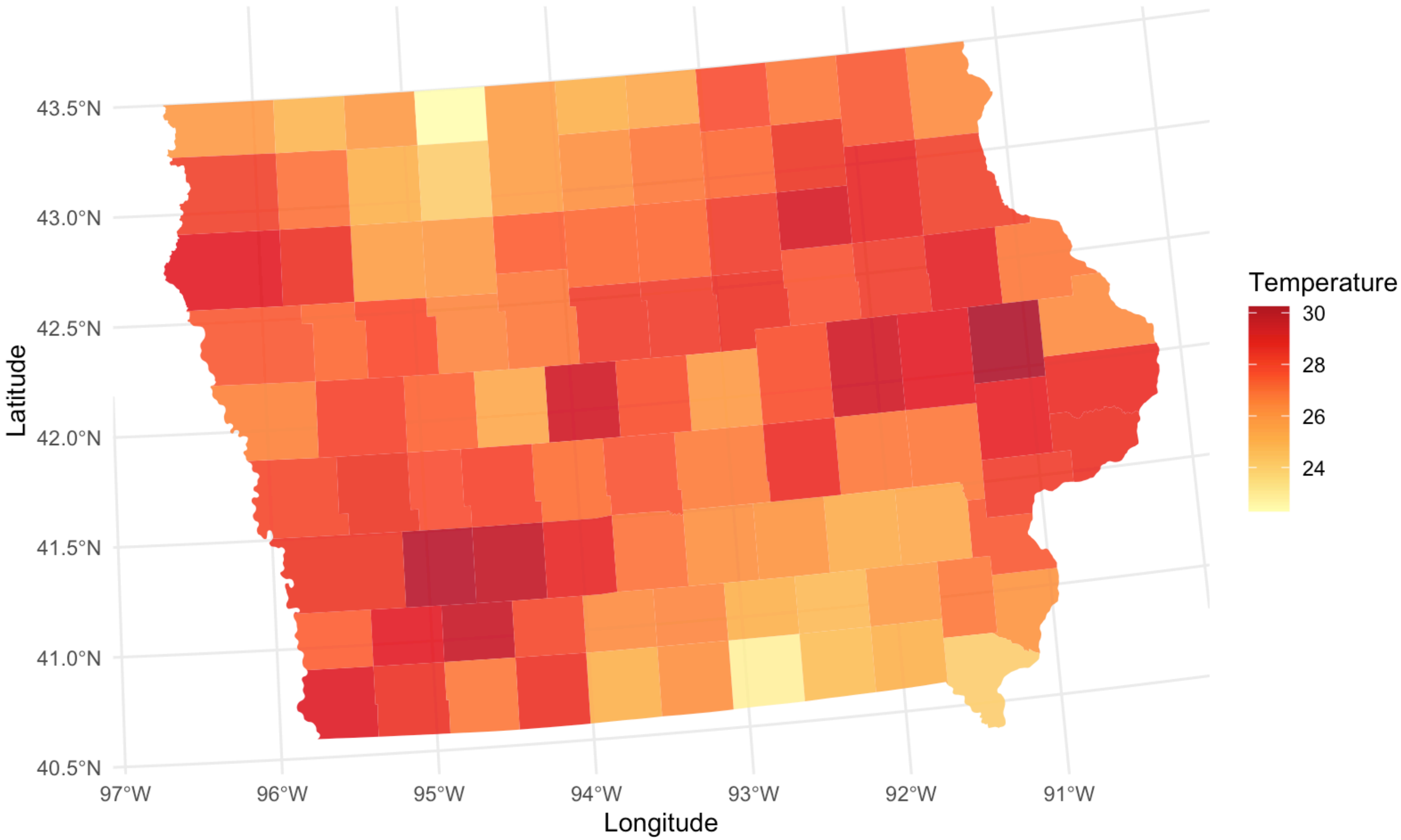
Estimate the county mean

- Visualising an estimate, such as a mean, can make trends easier to see
  - This estimate has a standard error, but we rarely integrate it into our plots

▶ Code

county_name	temp_mean	temp_se	n
Adair County	29.7	0.907	6
Adams County	29.6	1.003	9
Allamakee County	26.3	0.550	8
Appanoose County	22.8	0.831	14
Audubon County	27.6	0.893	11

Can you see the trend now?

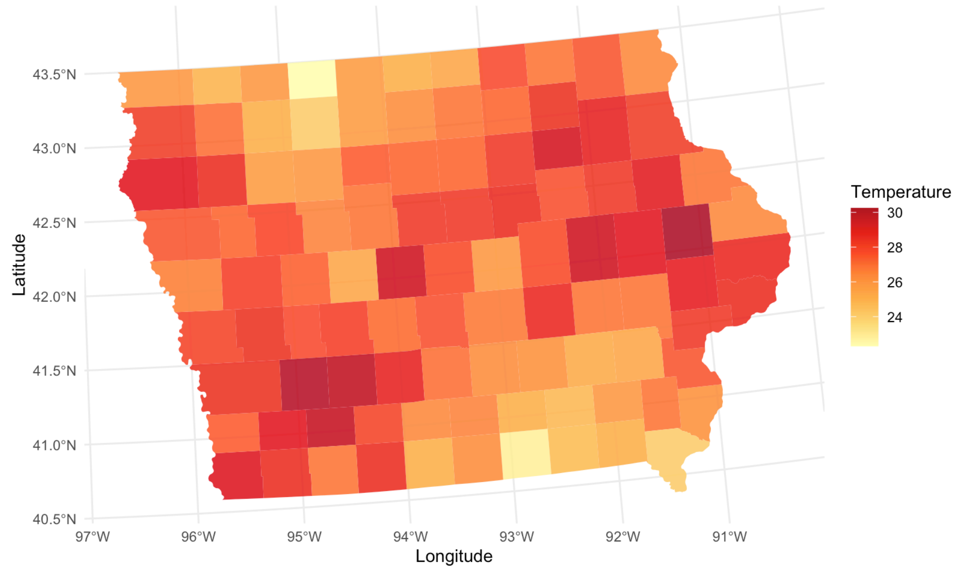


https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Common Map Visualisations

- Usually spatial data is shown using a choropleth map,
  - Choropleth maps shade an area according our statistic of interest
- We can also weight the area by another variable, such as sample size
  - e.g. cartograms, and bubble plots
- Does this plot follow the principles of signal suppression?
- Is there a noticeable difference in the way these plots convey signal?

Choropleth Map | Cartogram | Bubble Map

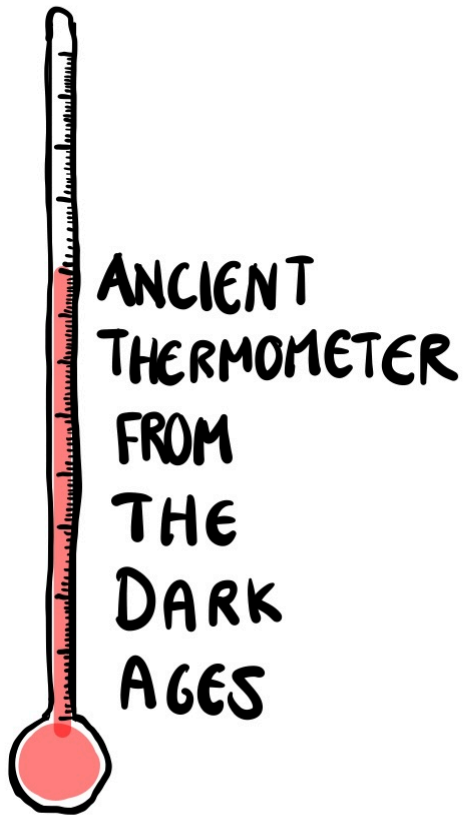


https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

But what if the error is worse?

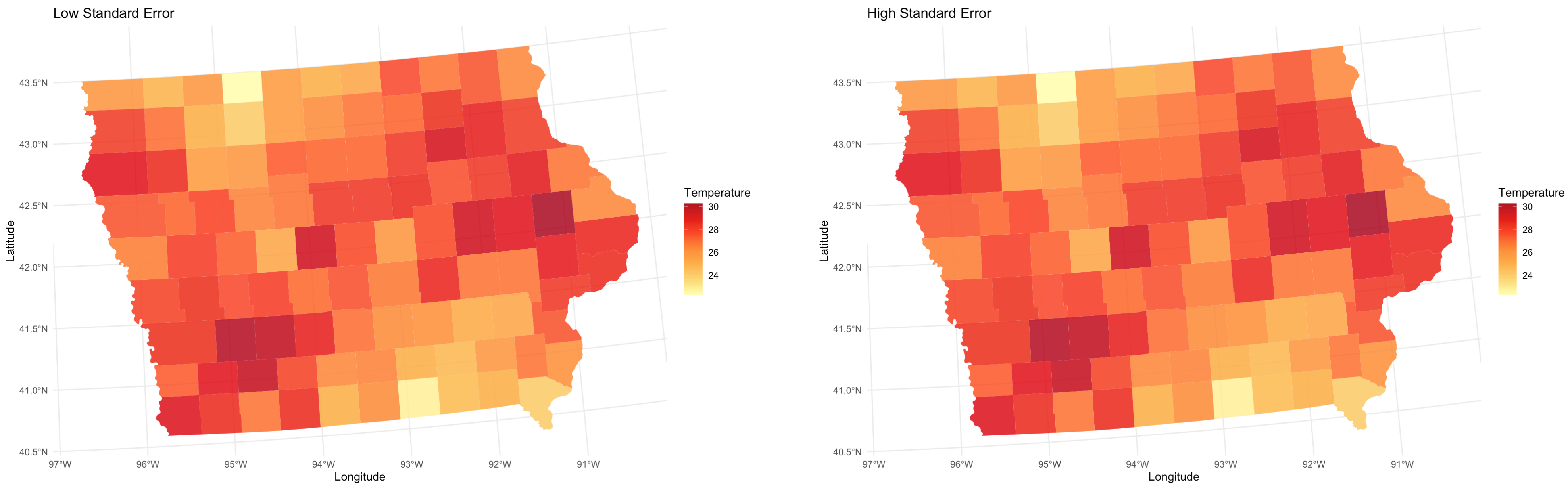
- It turns out the citizen scientists are using some pretty old tools.
- The standard error **could** be up to three times what we would estimate with our usual assumptions.
- We want to see both versions of the data so we can see the impact of this change.

county_name	temp_mean	low_se	high_se
Adair County	29.7	0.907	2.72
Adams County	29.6	1.003	3.01
Allamakee County	26.3	0.550	1.65



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Spot the difference



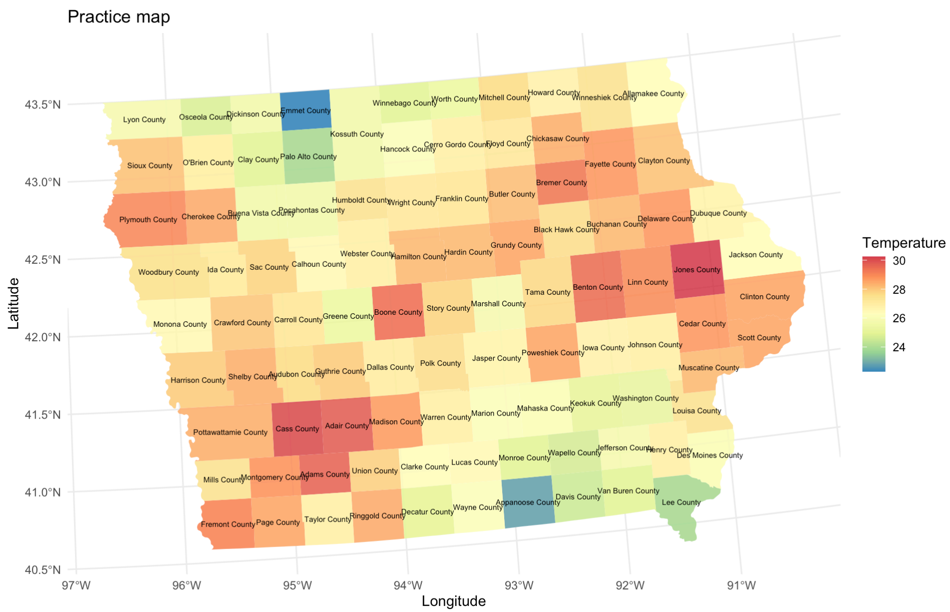
One of these maps was made using the estimate with the high standard error, the other was made with the estimate from the low standard error. Can you tell which is which?

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Exercise 1

Get comfortable working with spatial features yourself

- The citizen scientist data is available in the `ggdibbler` package as `toy_temp`.
- Go through the steps we worked through thus far in the tutorial, and get your estimate and standard error variables.
- Experiment changing the standard error to see if there is an impact on the plot.
- You can also try adding county names, changing the colour scale, and making other aesthetic changes to the map.



Approaches to Spatial Uncertainty

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Looking at current approaches

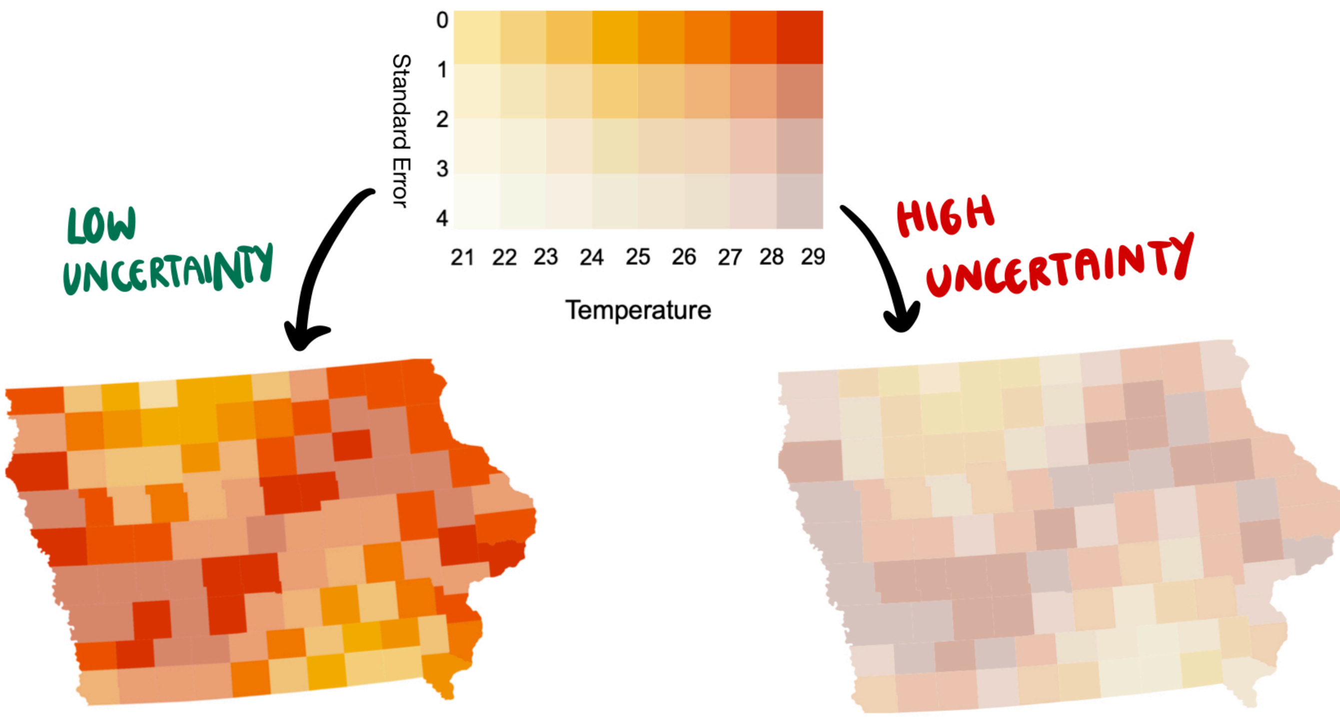
We are going to go through some uncertainty visualisation methods assess them on the signal suppression criteria.

> Remember, uncertainty visualisation should
>
> 1. Reinforce justified signals
>   *We want to trust the results*
> 2. Hide signals that are primarily noise
>   *We don't want to see something that isn't there*

Solution 1: add an axis for uncertainty (Bivar)

Questions to think about…

- Is there a visible difference between the high and low uncertainty cases?
- Is the trend still visible in the high uncertainty case?
- Is this approach accessible?
  - Is colour a simple 3D space?
  - Can everyone see changes in saturation?

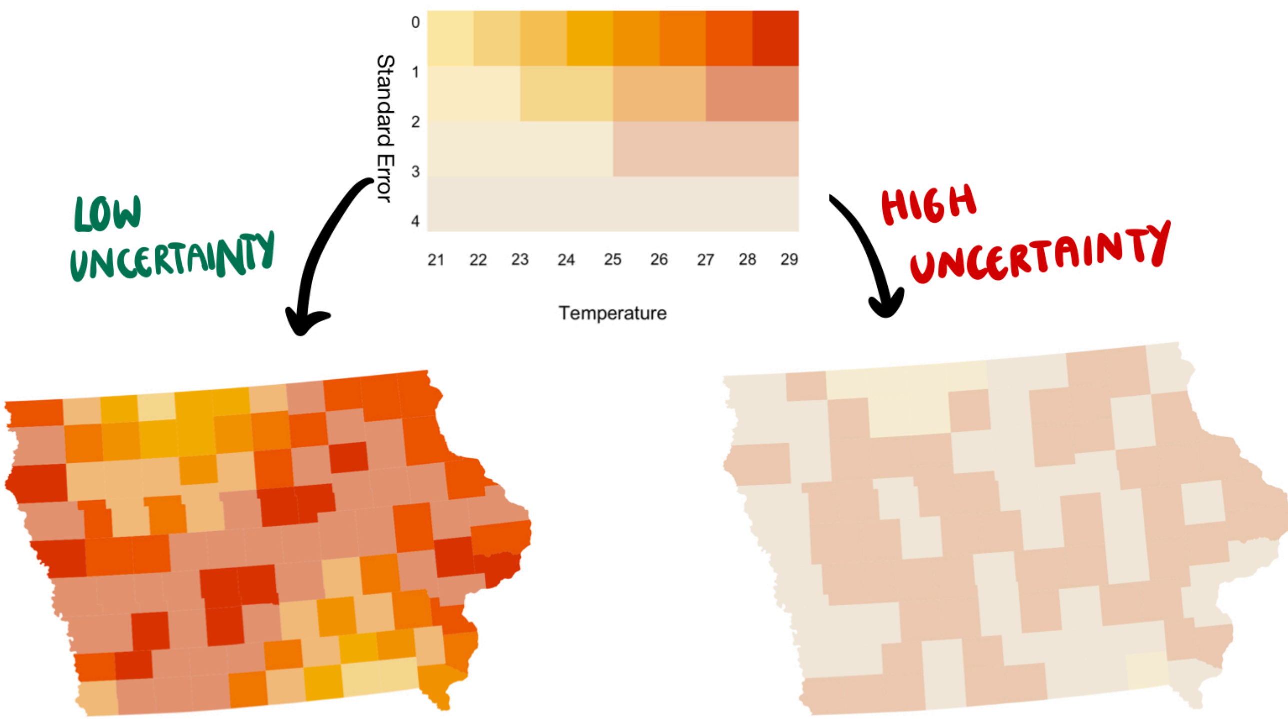


https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Solution 2: blend the colours together (VSUP)

Questions to think about…

- Is there a visible difference between the high and low uncertainty cases?
- Is the trend still visible in the high uncertainty case?
- Is this approach accessible?
- At what level of uncertainty should you blend two colours together?

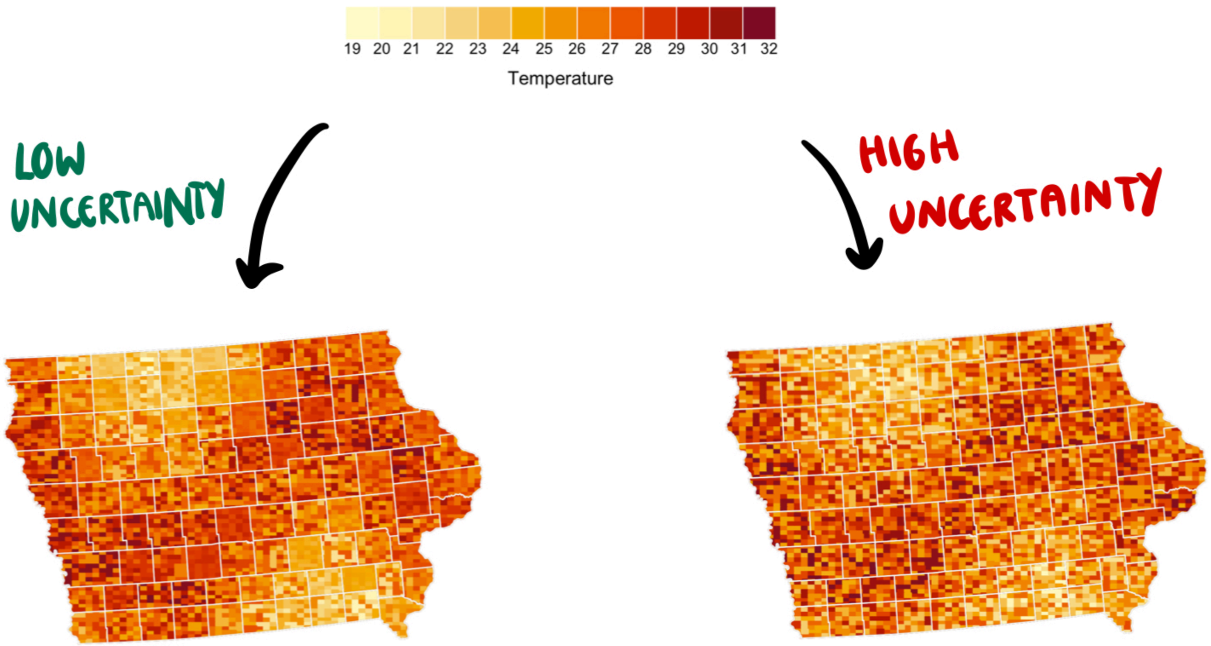


https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Solution 3: simulate a sample (pixel)

Questions to think about…

- Is there a visible difference between the high and low uncertainty cases?
- Is the trend still visible in the high uncertainty case?
- Is this approach accessible?
- What has replaced the manual colour blending in this approach?



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Popular R packages

- `ggdibbler`
  - Data: distribution from distributional
  - Maps: pixel map, other non-spatial maps
- `Vizumap`
  - Data: Depends on the plot. Can take a distribution as a q function (pixel), or an estimate and standard error as two variables (bivar/VSUP and glyph)
  - Maps: Bivar/VSUP, pixel, glyph
- `biscale`
  - Data: Estimate and standard error as two variables
  - Maps: Bivar/VSUP map

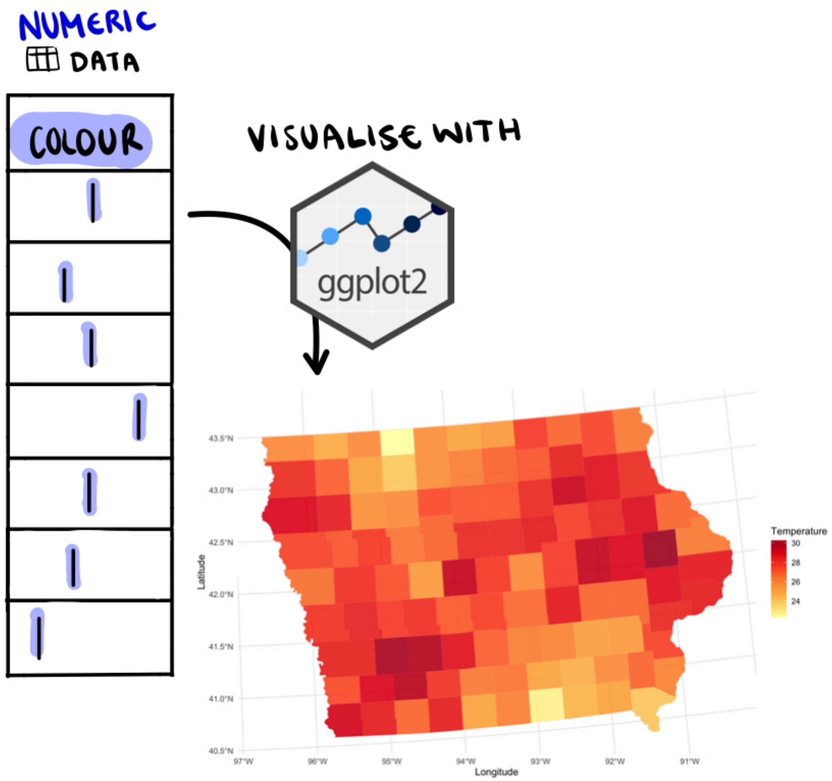
Making a Pixel Map with `ggdibbler`

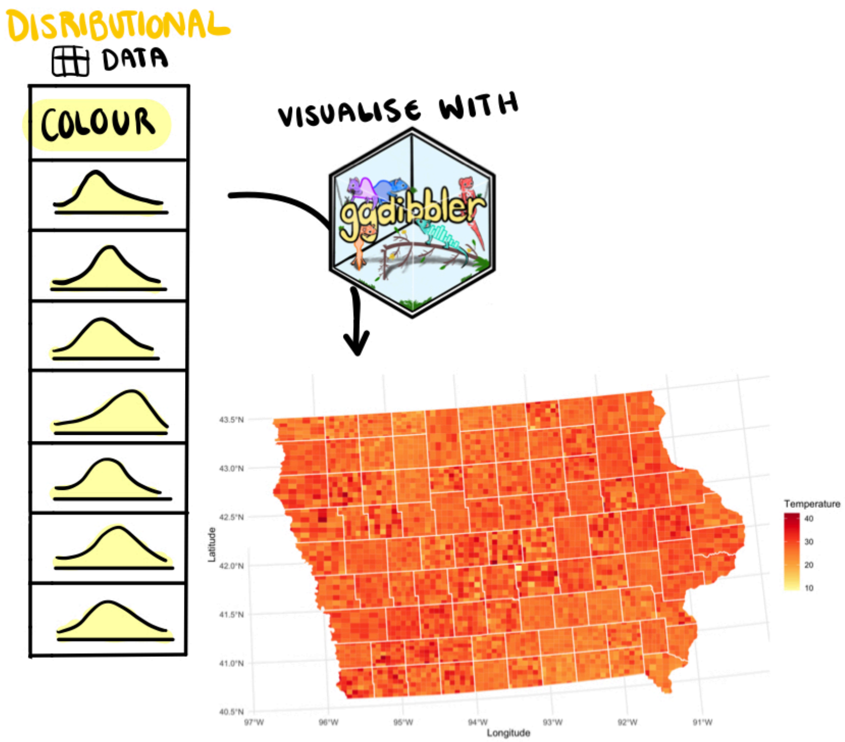
https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

A `ggdibbler` example

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

`distributional` and `ggdibbler` ecosystem





https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

Using distributional

- Expressing an estimate as a random variable using `distributional` makes the uncertainty more explicit
  - We will make a new variable, `temp_dist` that contains the sampling distribution of `temp_mean`
  - *Note: distributional uses standard deviation, but prints the variance.*

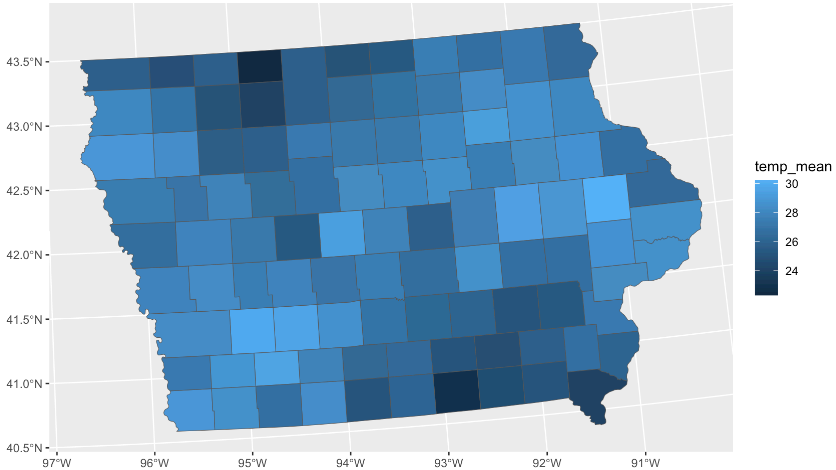
▶ Code

county_name	temp_mean	temp_se	temp_dist	n
Adair County	29.7	0.907	N(30, 0.82)	6
Adams County	29.6	1.003	N(30, 1)	9
Allamakee County	26.3	0.550	N(26, 0.3)	8
Appanoose County	22.8	0.831	N(23, 0.69)	14
Audubon County	27.6	0.893	N(28, 0.8)	11

Comparing ggplot to ggdibbler

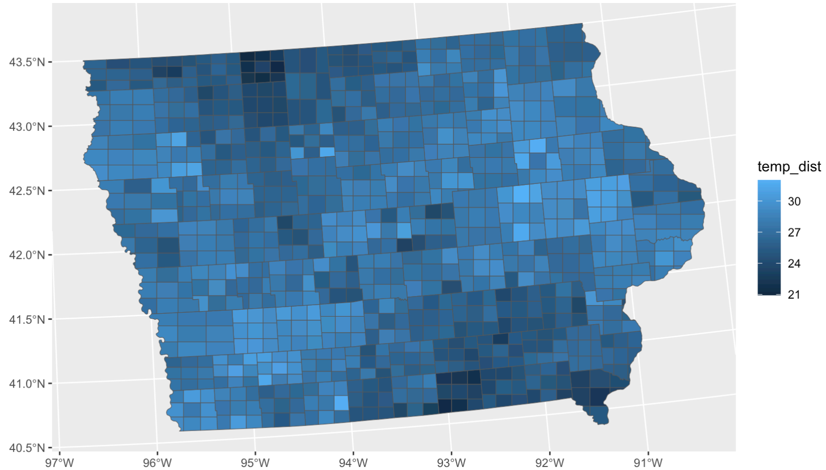
ggplot code

```r
toy_temp_mean |>
  ggplot() +
  geom_sf(aes(geometry = county_geometry,
                        fill=temp_mean))
```



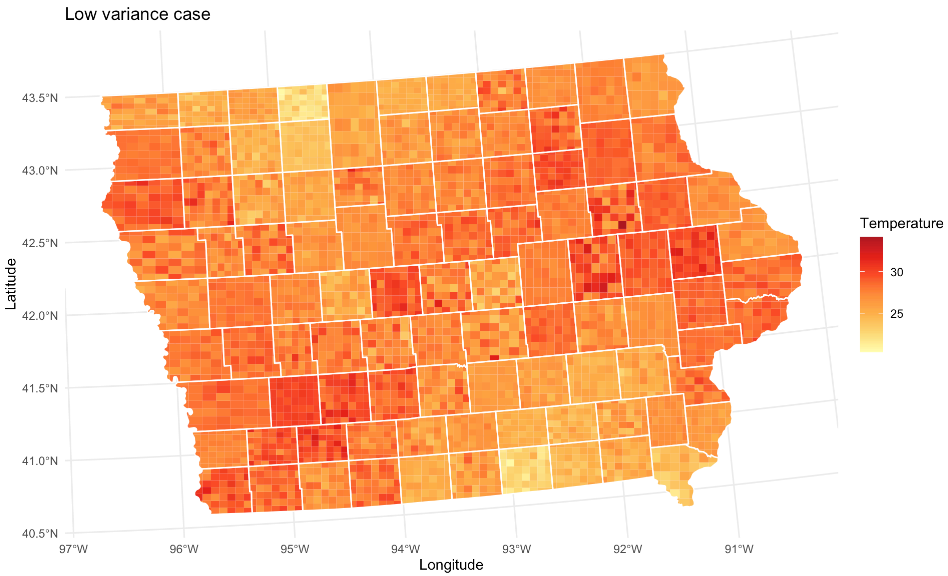
ggdibbler code

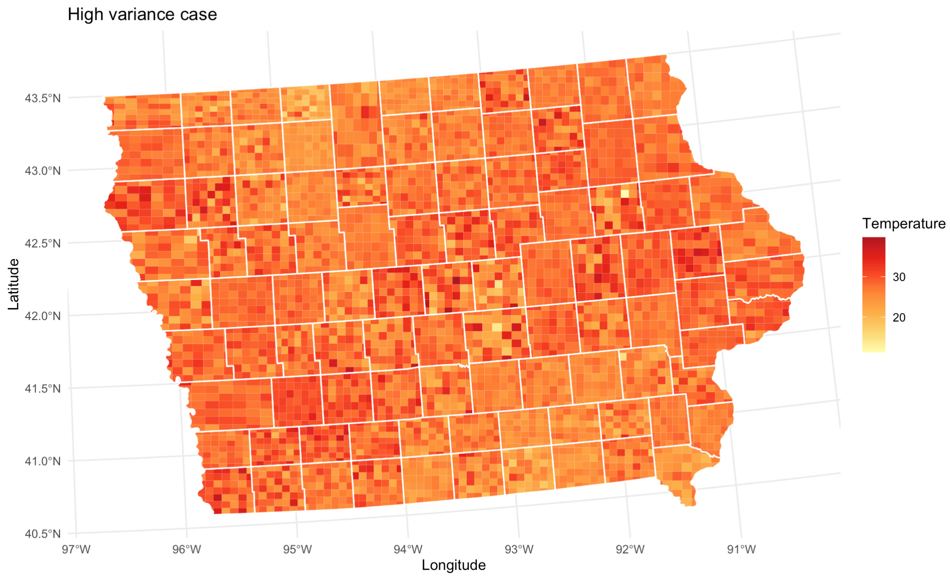
```r
toy_temp_mean |>
  ggplot() +
  geom_sf_sample(aes(geometry = county_geometry,
                        fill = temp_dist))
```



https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

High and low variance comparison with `ggdibbler`

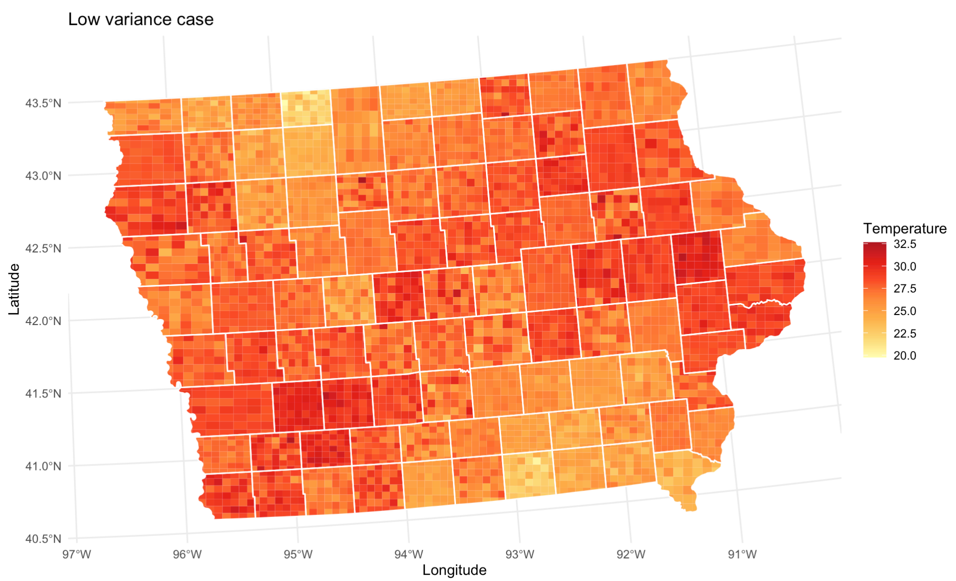


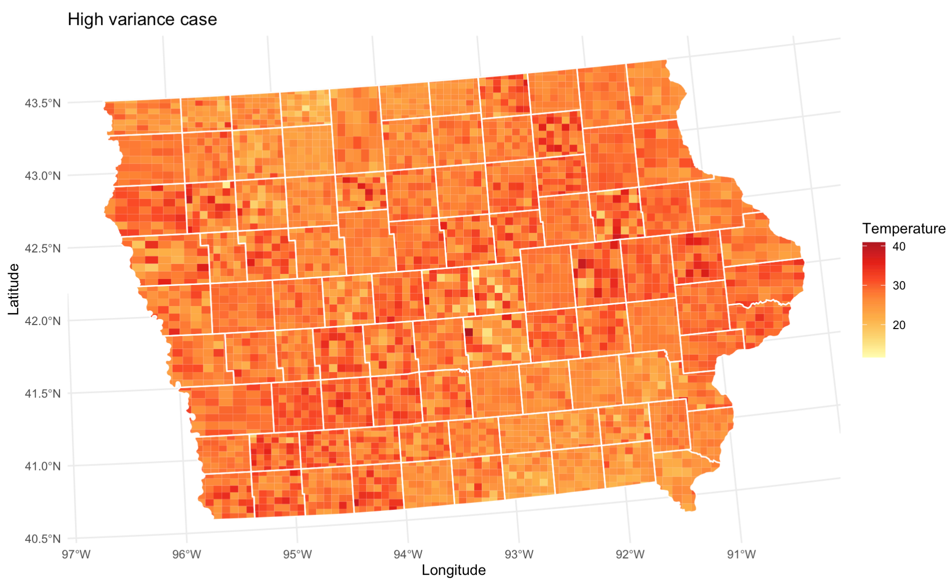


Remember, the plot is random

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

High and low variance comparison with `ggdibbler`





Remember, the plot is random

https://numbat-tutorials.github.io/tutorial_visualising_uncertainty/

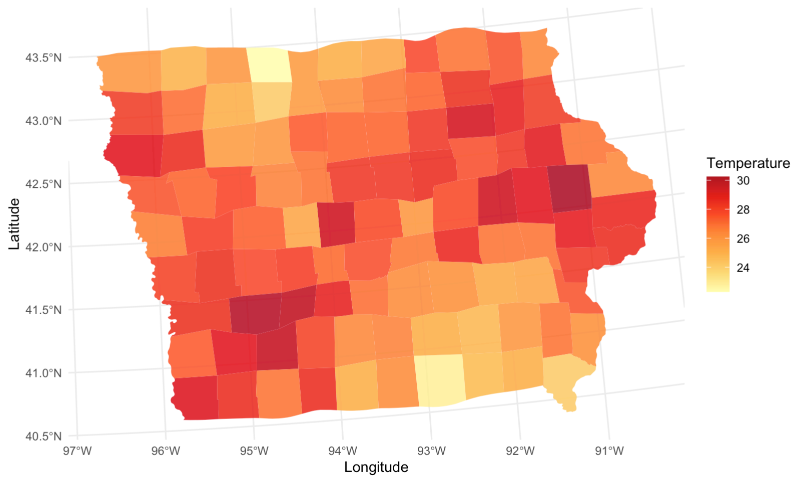
Exercise 2

Here is the code that was used to make the cartogram from earlier in the session. Using `distributional` and the standard error provided in `toy_temp_mean`, can you make a `ggdibbler` version of this plot?

10:00

Cartogram with no uncertainty | Check

▶ Code



Where to learn more

- Kay, M. ggdist: Visualizations of distributions and uncertainty
- Wilke, C. O. ggridges
- Pedersen, T. ggforce
- Hofmann, Follett, Majumder, Cook (2012) Graphical Tests for Power Comparison of Competing Designs
- nullabor package
- Spiegelhalter, D. (2017), Risk and uncertainty communication
- Correll, Moritz, Heer (2018) Value-Suppressing Uncertainty Palettes
- Hullman et al (2018) Imagining Replications
- Lucchesi, Kuhnert, Wikle Vizumap
- Distributional package
- ggdibbler package
- Kinkeldey, MacEachren, Riveiro, Schiewe (2017) Evaluating the effect of visually represented geodata uncertainty on decision-making
- Mason, Cook, Goodwin, Tanaka, VanderPlas (2024) The Noisy Work of Uncertainty Visualisation Research

End of session 2