



# BST 219

# Core Principles of Data Science

Lecture 18: Advanced Data Wrangling continued and maps  
October 31, 2024

# Recipes of the Day!

## Diwali Sweets Recipes



### DIWALI SWEETS

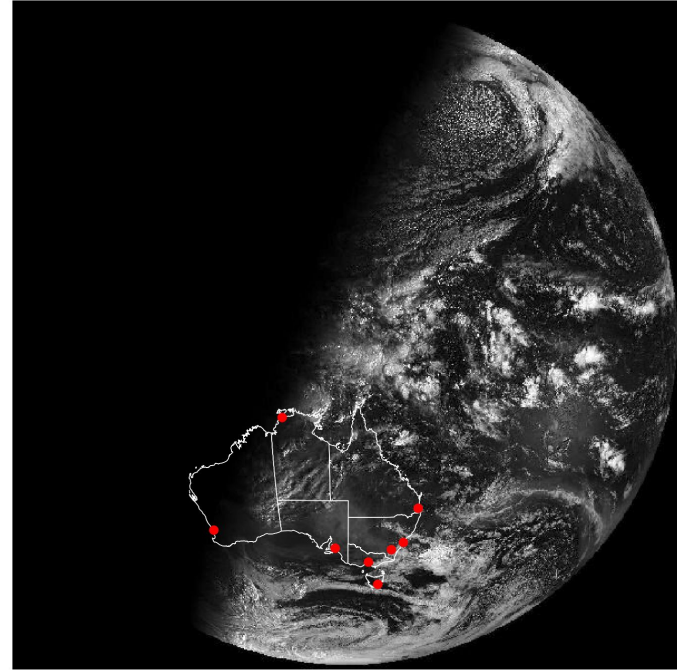
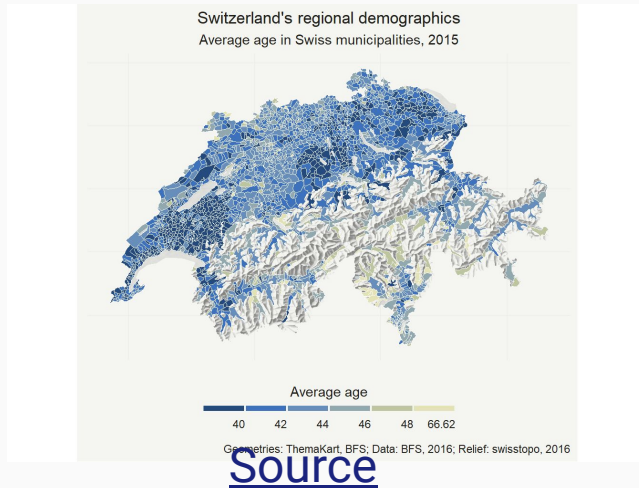


## Dia de los Muertos Recipes



# Agenda

- Announcements
  - Lab this week!
  - Homework 3 is due 11/8
  - Midterm 11/8 - 11/17
- Finish dates and times
- TableOne package
- Maps



Source

# Special Announcements

- Heather's **11/5** office hour will be moved to **12-1pm**
- Heather's **11/12** office hour will be moved to **12-1pm**
- The 11/14 lecture will be moved to **11/13, 12:30-2pm via Zoom**
- Lecture on **11/26** will be held via **Zoom**
- Heather's office hour on **11/26** will be **Zoom only**



# Coding Question of the Day!

Using the `gapminder` dataset, first remove all rows with missing `gdp` data (code included below). Then, create 2 separate data frames: one containing data just from the year 1975 and one containing data just from the year 1995. Then create one dataset containing data about countries that appear in both the 1975 and 1995 datasets, and one dataset containing data from countries that only appear in the 1995 dataset.

**Bonus challenge:** Create a bar plot summarizing the number of countries on each continent that are in the dataset containing data about countries that only appear in the 1995 dataset.

Make sure to run  
this code first

```
library(dplyr)
library(dslabs)
library(ggplot2)
library(tidyr)

data("gapminder")

# Drop all rows with NA (missing data) for gdp
gapminder <- gapminder %>% drop_na(gdp)
```