

Students alcohol consumption

DNDS 6002: DATA AND NETWORK VISUALIZATION

Name : Hao CUI
December 4, 2018

CONTENTS

Motivation

1

Dataset overview

2

Data visualization

3

Project 2

4

Motivation

- Visualize students academic performance by age and whether they are heavy drinkers.
- Visualize students alcohol consumption on workdays and weekends by age and gender.
- We will explore and arrive at the most appropriate visualizations step by step.

Dataset overview

Downloaded from Kaggle
Dimension: 649×33

Important variables:

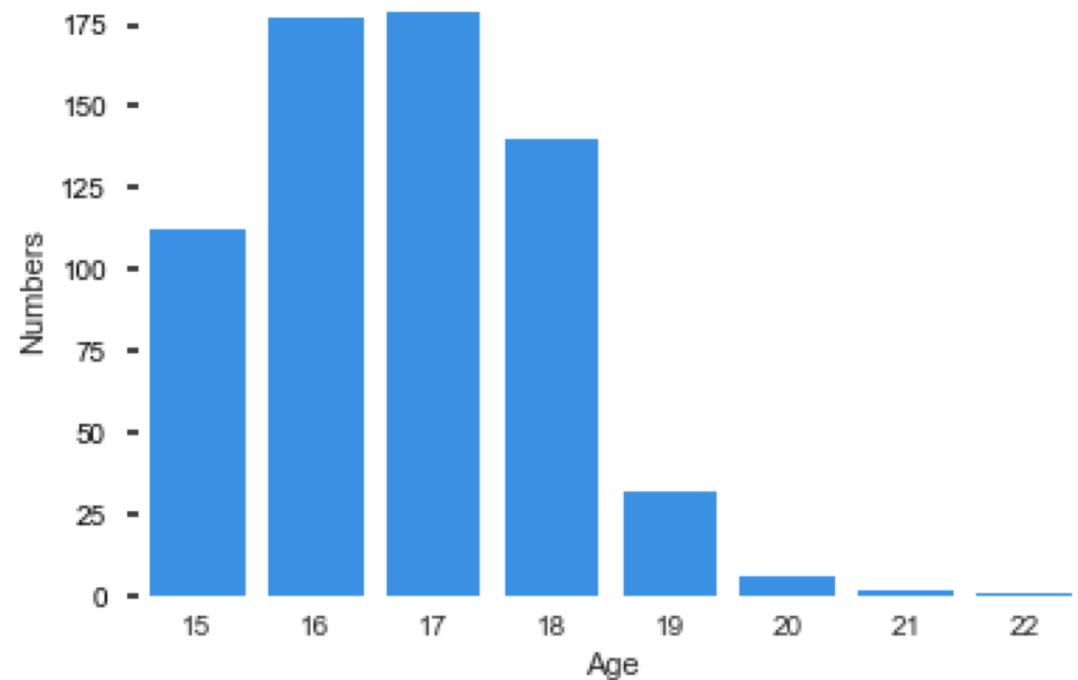
Dalc, Walc - workdays and weekends alcohol consumption (numeric: from 1 to 5)

Create variable Talc - $Talc = (5 \times Dalc + 2 \times Walc) / 7$

Create a variable High - (Binary: yes and no)
Choose a threshold 2, over 2 is high

Gender ratio F : M = 0.59 : 0.41

Students age distribution



Dataset visualization



Since the gender ratio is not 1:1,
set scale="area" in the violinplot

Disadvantage:

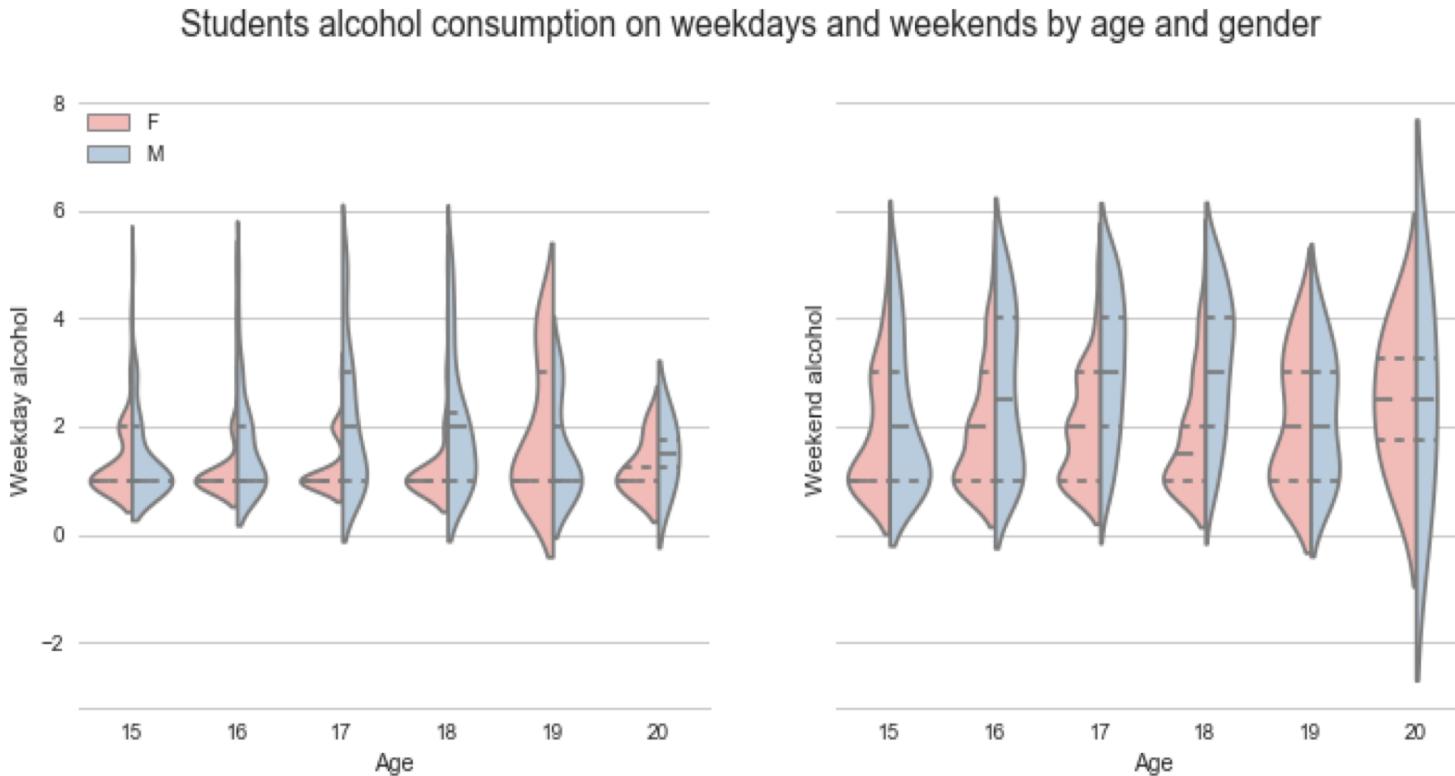
- Boxplots inside violinplots
can't represent female and
male students individually.

Dataset visualization



Use `inner="quartile"` inside the violinplots to show the quartiles of both female and male students.

Dataset visualization



Splitting violinplots by gender

Disadvantages:

- Some quartiles are not shown/easily recognized
- The tail of the age 20 violinplot doesn't seem to be correct (Violins for relatively small samples might look misleadingly smooth.)

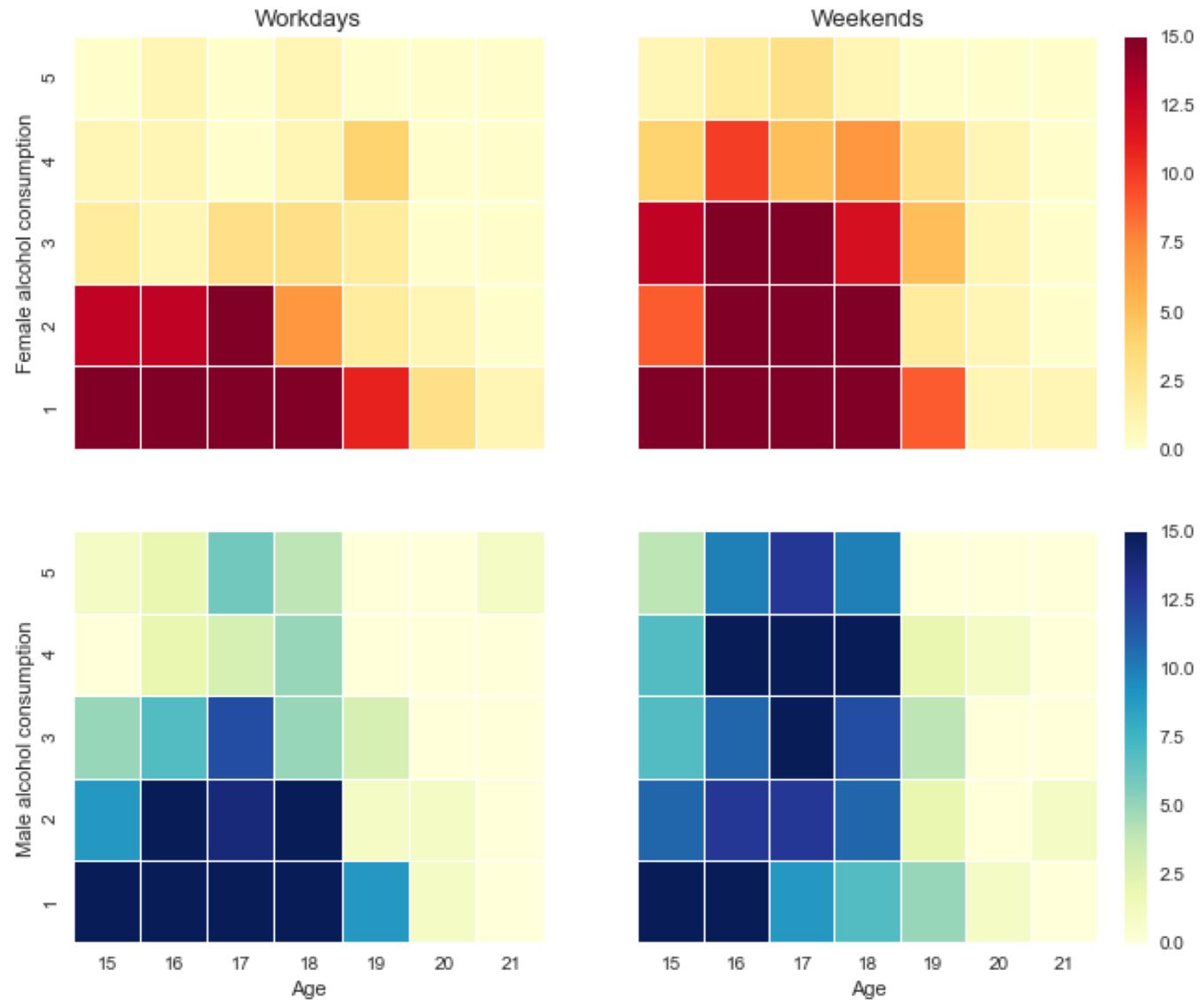
Dataset visualization

Version before the final

Disadvantages:

- Workdays and weekends separation cause confusion
- Labels of y axis contains many words

Students alcohol consumption on workdays and weekends by gender and age



Dataset visualization

Advantages:

Convey information:

- Make comparison of workdays and weekends alcohol consumption by gender and age

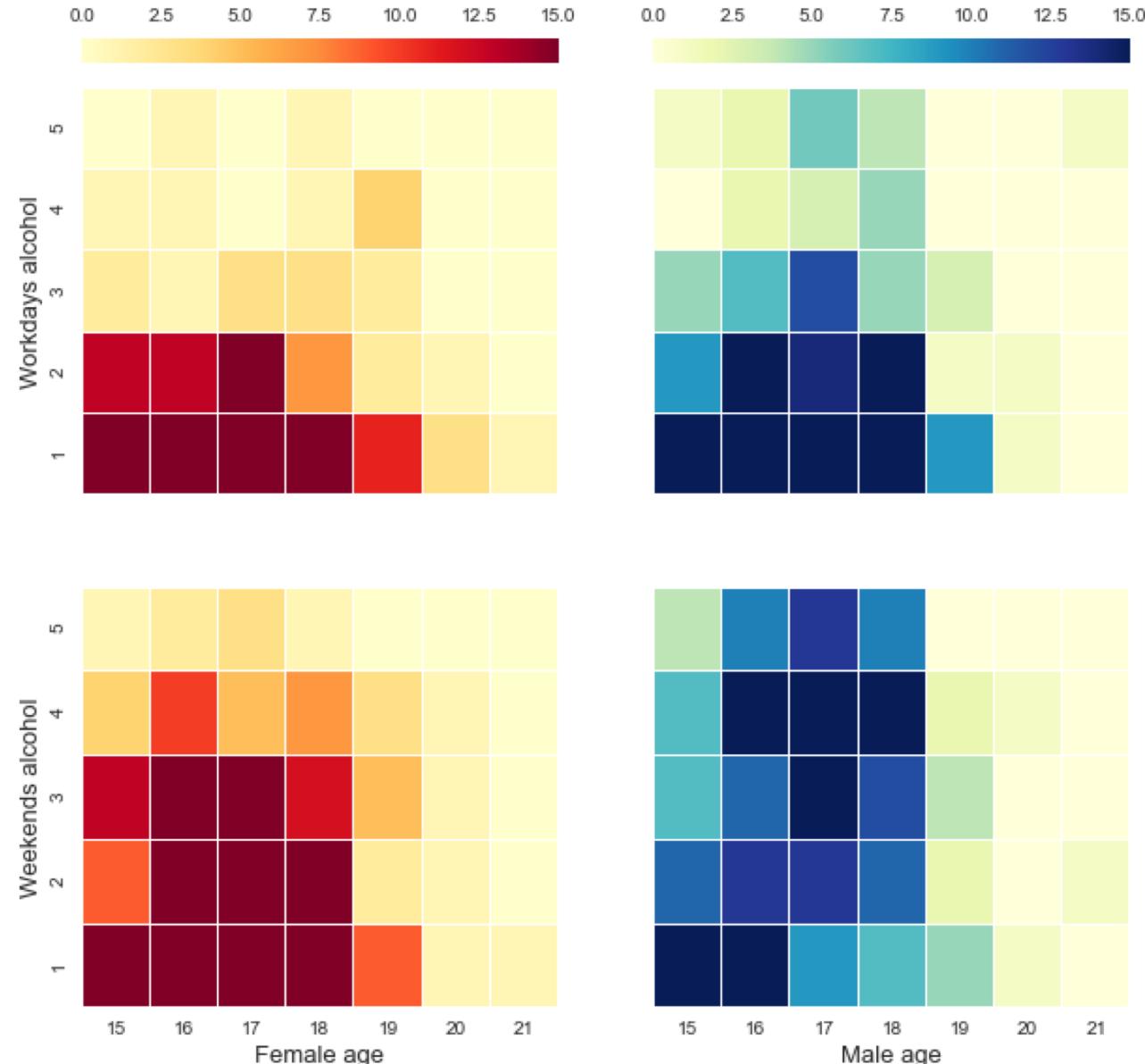
Reduce redundancy

- Two horizontal subplots share y axis
- Two vertical subplots share x axis and colormap

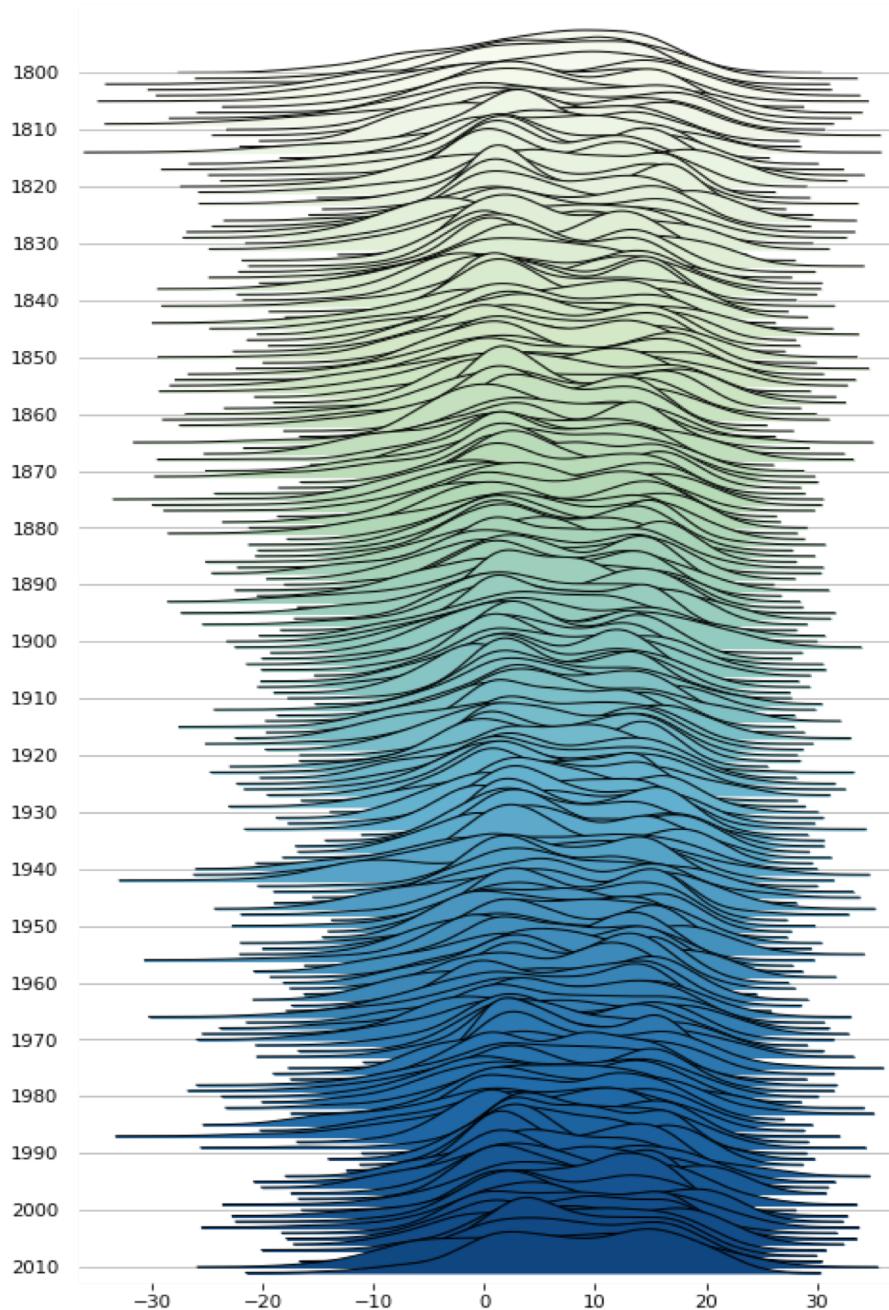
Appropriate colormaps:

- Red for female, blue for male
- Dark for high and light for low

Students alcohol consumption on workdays and weekends by gender and age



Stockholm yearly temperatures from 1800-2011



Project 2 – joyplot/ridgeline plot

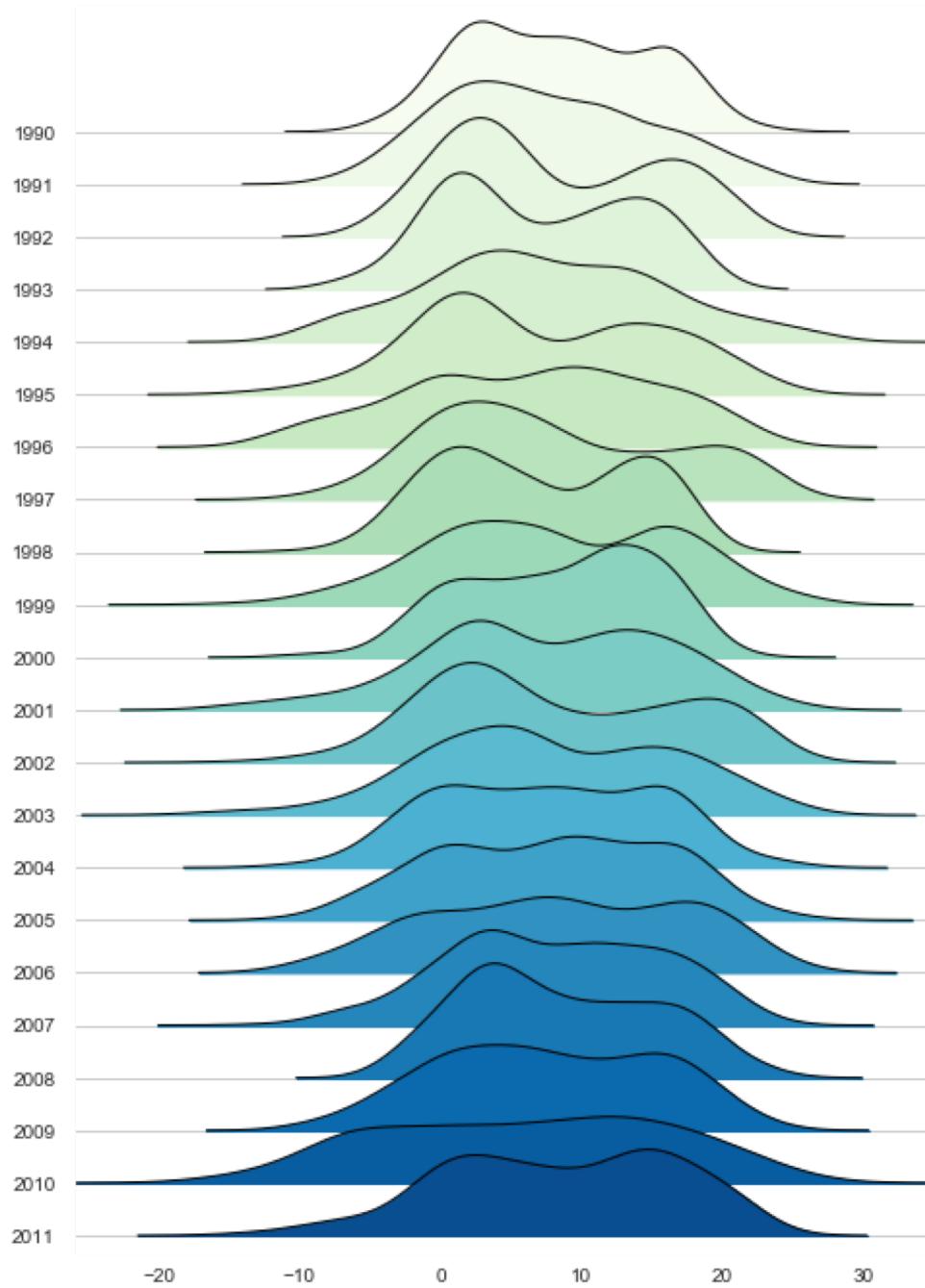
Dataset description:

- Dataset taken from data visualization class04
- Stockholm temperatures from 1800 to 2011
- Dimension: 77431×4

Disadvantages:

- Distribution areas are greatly overlapping, former years distributions are hidden behind

Stockholm yearly temperatures from 1990-2011



Advantages:

- Distribution areas are slightly overlapping, easy to see the temperature distributions of each year.

Discussions :

- Can be quite useful for visualizing changes in distributions over time or space
- Especially interesting when the number of groups to represent is high
- Be careful with overlapping data
- Suitable colormap is important

Thanks for watching!

Name : Hao CUI
December 4, 2018