

Review 1

Abinash Satapathy

Reg. no.: 16BCE0081

CSE3020 – Data Visualisation

Slot: C1

Faculty: Lydia Jane G

Visualisation of **Global Temperature** using R

1. Dataset details

- URL: <https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data/downloads/climate-change-earth-surface-temperature-data.zip/2>
- No. of attributes = 7
No. of rows = 239, 178
- Attributes definition:

Date	The date whose temperature is included in the dataset
Average temperature	The mean temperature for that day
Uncertainty in calculation	The error amount in calculation of the temperature
City	Place whose temperature is taken into consideration
Country	The country where the place is situated
Latitude	The angular distance of a place north or south of the earth's equator
longitude	The angular distance of a place east or west of the Greenwich meridian

2. Data Abstraction

Date	Quantitative
Average temperature	Quantitative
Uncertainty in calculation	Quantitative
City	Qualitative (nominal)
Country	Qualitative (nominal)
Latitude	Quantitative
longitude	Quantitative

3. Task Abstraction

- Potential questions that can be asked by looking at the dataset

- (i) What is the average temperature of a month of all the cities of a country during a particular year range?
- (ii) What is the year-wise average of different cities of a country?
- (iii) What are the peak temperatures of a set of cities?
- (iv) What is the temperature rise/fall over a wide range of years (like 50 to 100 years range)?
- (v) Which country has the highest mean temperature for a specific month?
- (vi) Which country has the highest range of mean temperatures?
- (vii) What is the range of latitudes where the mean temperature is greater than some 'x' value?
- (viii) Is climate change really happening?
- (ix) What is the annual increase/decrease of the mean temperature of a country?
- (x) What is the difference pattern of temperature differences between mean temperature of a city and human body, where 37°C is the normal temperature of a human body?

b. Attributes required for answering the above questions and others as well

- (i) The Average Temperature attribute is the most important numerical data in the dataset. So, usually all the queries will include the temperature data.
- (ii) The city attribute is important for questions targeted to a particular city or a set of cities.
- (iii) The country attribute is of equal importance as the city, as it helps in getting patterns on a global scale.

The actions that would be performed mostly are:

- Plotting the temperature
- Calculating the range of mean temperatures by Highest – Lowest formula
- Standard deviation formula might be used in case a query regarding that is asked.

$$SD = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

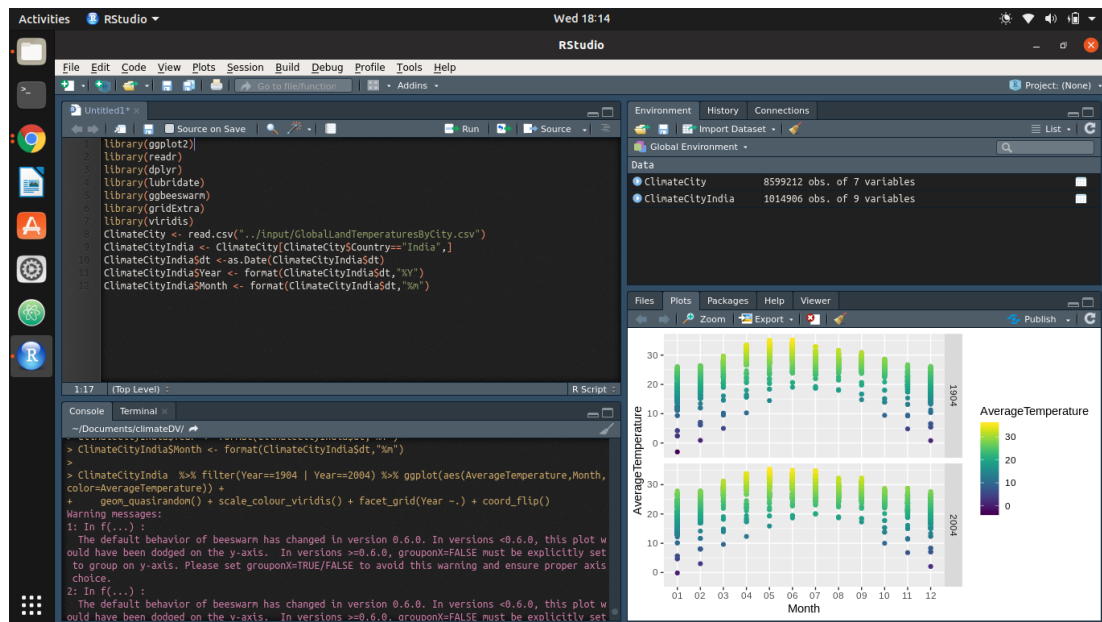
4. Encoding

a. Appropriate encoding techniques

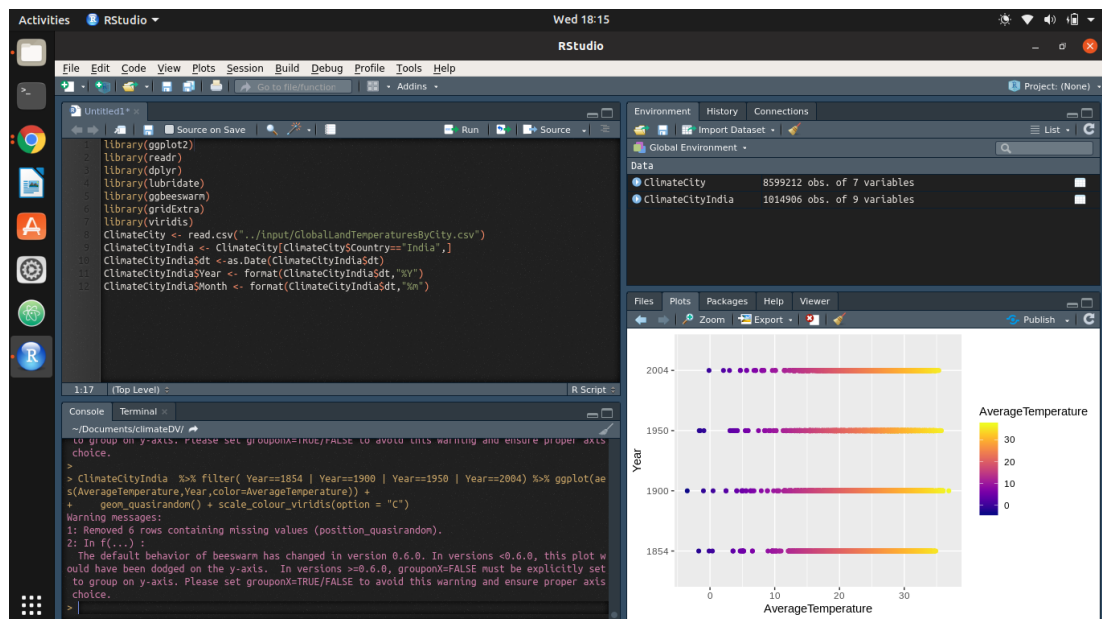
- (i) Jitter plot
- (ii) Line graph
- (iii) Heat map
- (iv) Density plot
- (v) Scatterplot

b. Example graphs for analysis of certain questions

- (i) ClimateCityIndia %>% filter(Year==1904 | Year==2004) %>%
ggplot(aes(AverageTemperature,Month,color=AverageTemperature)) +
geom_quasirandom() + scale_colour_viridis() + facet_grid(Year ~.) + coord_flip()

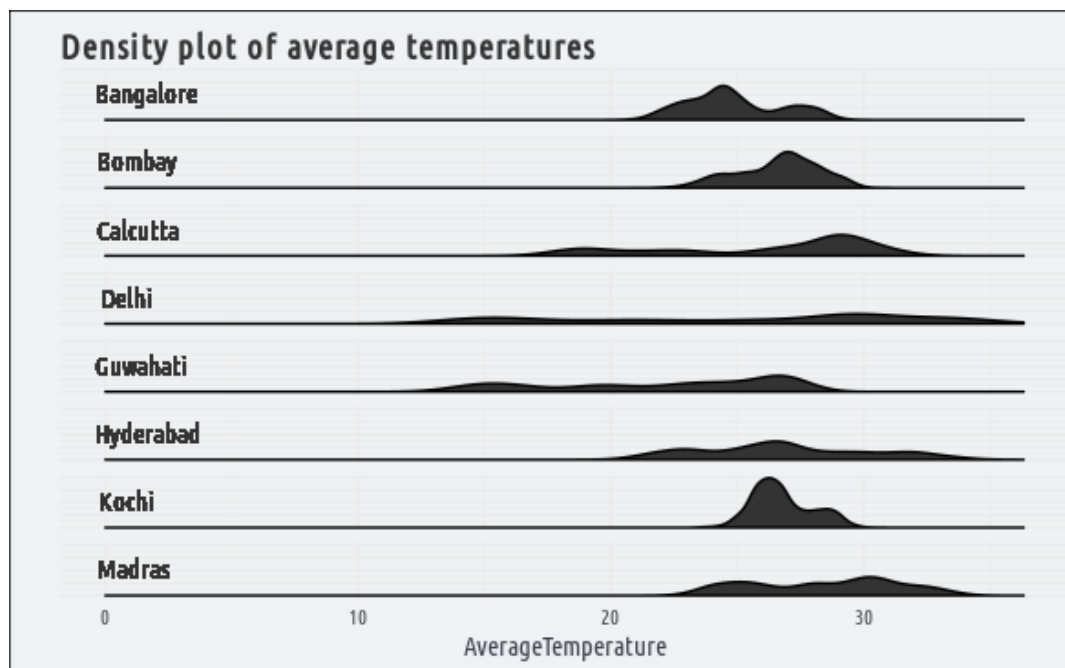


- (ii) ClimateCityIndia %>% filter(Year==1854 | Year==1900 | Year==1950 | Year==2004)
%>% ggplot(aes(AverageTemperature,Year,color=AverageTemperature)) +
geom_quasirandom() + scale_colour_viridis(option = "C")



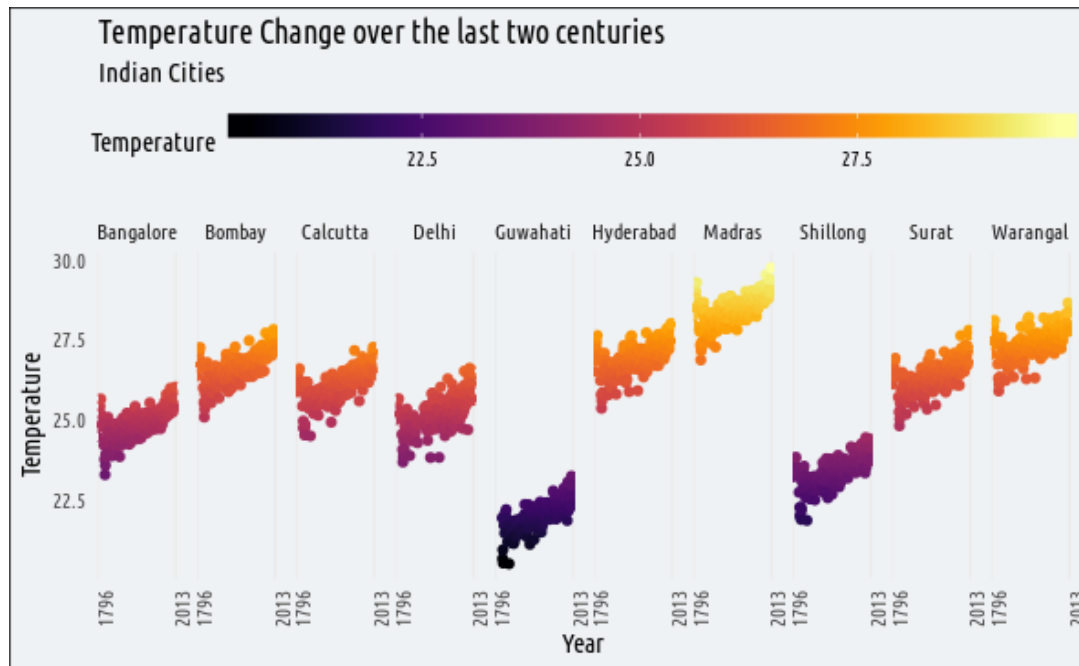
- (iii) Major_Cities<- c("Madras","Hyderabad","Bombay","Bangalore","Delhi",
,"Calcutta","Guwahati","Kochi")
df <- ClimateCityIndia %>% filter(City %in% Major_Cities)
ggplot(df,aes(AverageTemperature,label=paste(City,"")))+ geom_density(fill="gray20") +
facet_grid(City ~.) + ylab("") + geom_text(aes(0,0.2),family="Ubuntu")

```
Condensed",fontface="plain",hjust=0.1,color="gray20") + theme_minimal() +
ggtitle("Density plot of average temperatures") + theme(text = element_text(family =
'Ubuntu Condensed', face = 'plain', color = '#3A3F4A'), axis.text.y = element_blank(),
axis.ticks.y = element_blank(), axis.line.y = element_blank(), strip.background =
element_blank(), strip.text.y = element_blank(), axis.line.x = element_blank(),
plot.background = element_rect(fill = "#EFF2F4"), plot.title = element_text(size = 14, face =
"bold", colour = "gray20", vjust = -1))
```

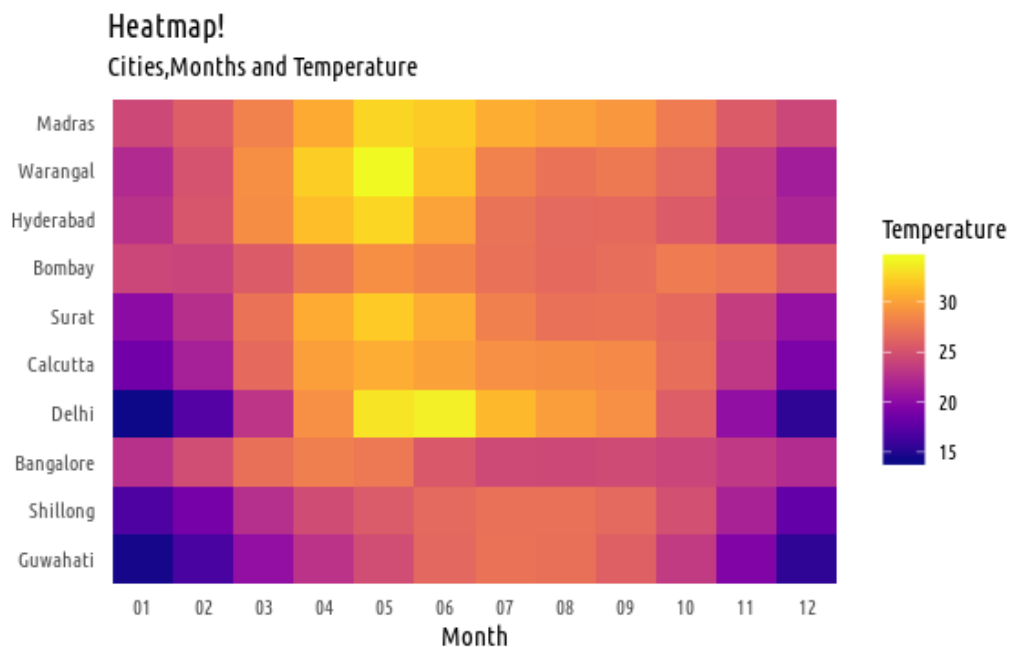


(iv) `Major_Cities<- c("Madras","Hyderabad", "Bombay", "Bangalore" , "Delhi" ,
,"Calcutta","Guwahati", "Shillong", "Surat", "Warangal")`

```
ClimateCityIndia %>% filter(City %in% Major_Cities) %>% group_by(City,Year) %>%
summarise(avg_temp= mean(AverageTemperature)) %>%
ggplot(aes(Year,avg_temp,color=avg_temp)) + geom_point() + scale_x_discrete( breaks =
c(1796,2013)) + theme_minimal(base_family = "Ubuntu Condensed") +
scale_color_viridis(option="B") + theme(axis.text.x = element_text(angle = 90, hjust = 1),
panel.spacing = unit(0.6, "lines"), legend.position = "top", legend.text=element_text(size=8),
panel.grid.major.y = element_blank(), panel.grid.minor.y = element_blank(), axis.ticks.y =
element_blank(), axis.ticks.x = element_blank(), plot.background = element_rect(fill =
"#EFF2F4"), legend.key.height = unit(10, "pt"), legend.key.width = unit(70, "pt")) +
ggtitle("Temperature Change over the last two centuries",subtitle = "Indian Cities") +
ylab("Temperature")+labs(color="Temperature") + facet_grid(~City )
```



- (v) `ClimateCityIndia %>% filter(City %in% Major_Cities) %>% group_by(City,Month) %>% summarise(avg_temp = mean(AverageTemperature, na.rm=T)) %>% ggplot(aes(Month, reorder(City, avg_temp), fill=avg_temp)) + geom_tile() + scale_fill_viridis(option = "C") + theme_minimal(base_family = "Ubuntu Condensed")+ labs(fill= "Temperature ")+ ylab("") + theme(panel.grid= element_blank()) + ggtitle("Heatmap!", subtitle = "Cities,Months and Temperature")`



- (vi) Average Temperature in India

Average Temperature in India

1796 to 2013

