

DSA2101 Problems for Fun!

2024-11-18

Disclaimer:

This assignment is provided solely for practice purposes (and fun of course). The questions included are designed to help you test and develop your skills with the **tidyverse**, especially ggplot in R. The problems may vary in difficulty, and they are not indicative of the types or difficulty level of questions you might encounter in any formal exams. If you encounter any difficulties or have questions regarding the problems, feel free to message me for assistance.

Good luck, and happy coding!

1. Airquality Dataset

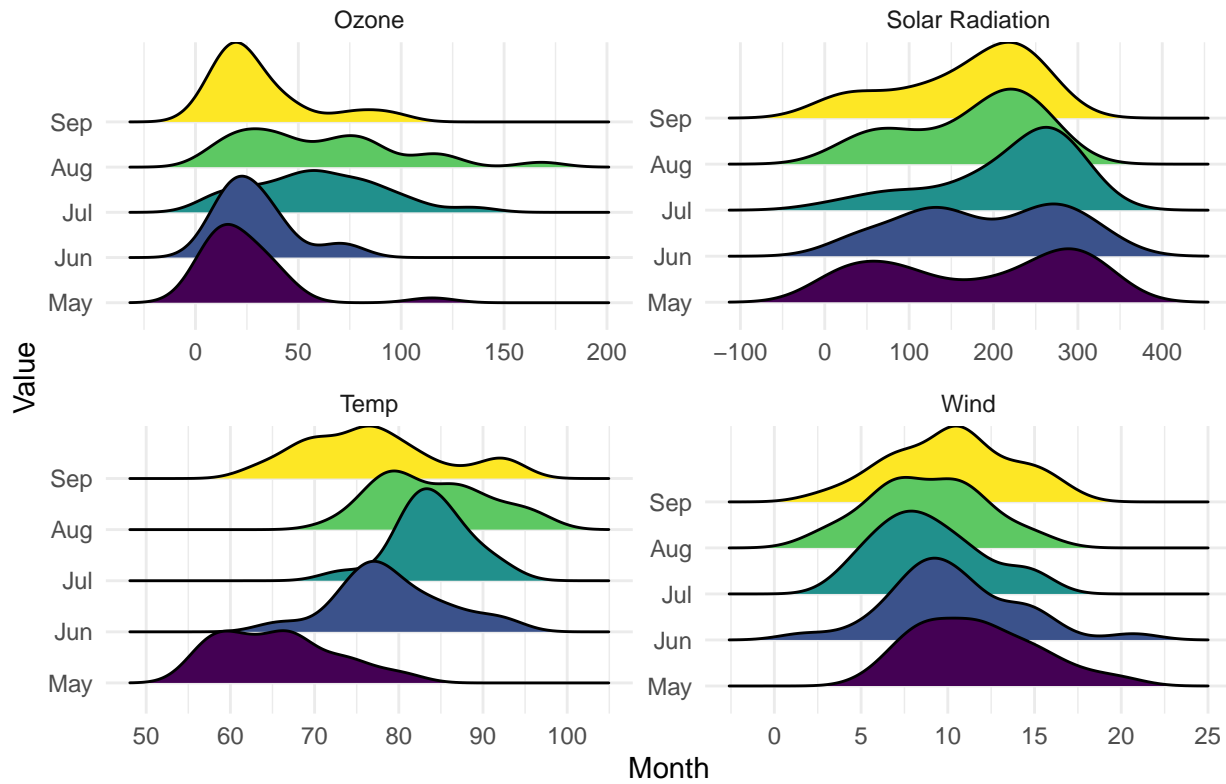
The airquality dataset provides valuable data on environmental conditions in New York, including measurements of Ozone, Solar Radiation, Wind, and Temperature recorded over several months. Your task is to perform an exploratory data analysis (EDA) and create compelling visualizations to uncover trends, patterns, and insights from the dataset. Create visualizations that showcase the variation in measurements (Ozone, Solar Radiation, Wind, and Temperature) across months.

```
# Load the Data  
data("airquality")  
head(airquality)
```

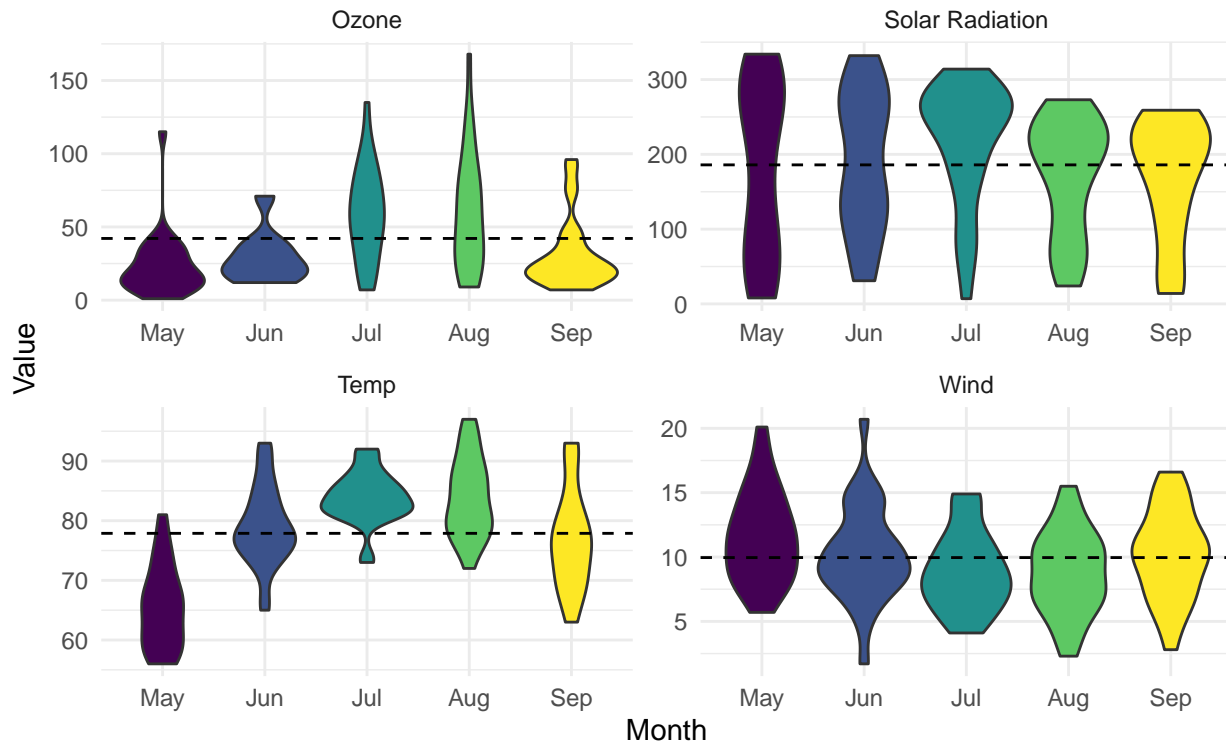
```
##   Ozone Solar.R Wind Temp Month Day  
## 1    41     190  7.4   67     5   1  
## 2    36     118  8.0   72     5   2  
## 3    12     149 12.6   74     5   3  
## 4    18     313 11.5   62     5   4  
## 5    NA      NA 14.3   56     5   5  
## 6    28      NA 14.9   66     5   6
```

Here are some samples:

Wind, Ozone, Solar and Temperature Levels Over Months in New York



Wind, Ozone, Solar and Temperature Levels Over Months in New York



Dashed line denotes mean.

2. Economics Dataset

The economics dataset in R provides U.S. economic indicators over time, including the number of unemployed individuals (unemploy) from 1967 onwards. Two major periods of economic downturn, the Dot-com Bust (2000–2002) and the Great Recession (2007–2009), are areas of interest in this analysis. Your goal is to show their impact on the US Economic Indicators

```
# Load the Data
data("economics")
head(economics)
```

```
## # A tibble: 6 x 6
##   date       pce    pop psavert uempmed unemploy
##   <date>    <dbl> <dbl> <dbl> <dbl>   <dbl>
## 1 1967-07-01 507. 198712  12.6   4.5    2944
## 2 1967-08-01 510. 198911  12.6   4.7    2945
## 3 1967-09-01 516. 199113  11.9   4.6    2958
## 4 1967-10-01 512. 199311  12.9   4.9    3143
## 5 1967-11-01 517. 199498  12.8   4.7    3066
## 6 1967-12-01 525. 199657  11.8   4.8    3018
```

Here is a sample:



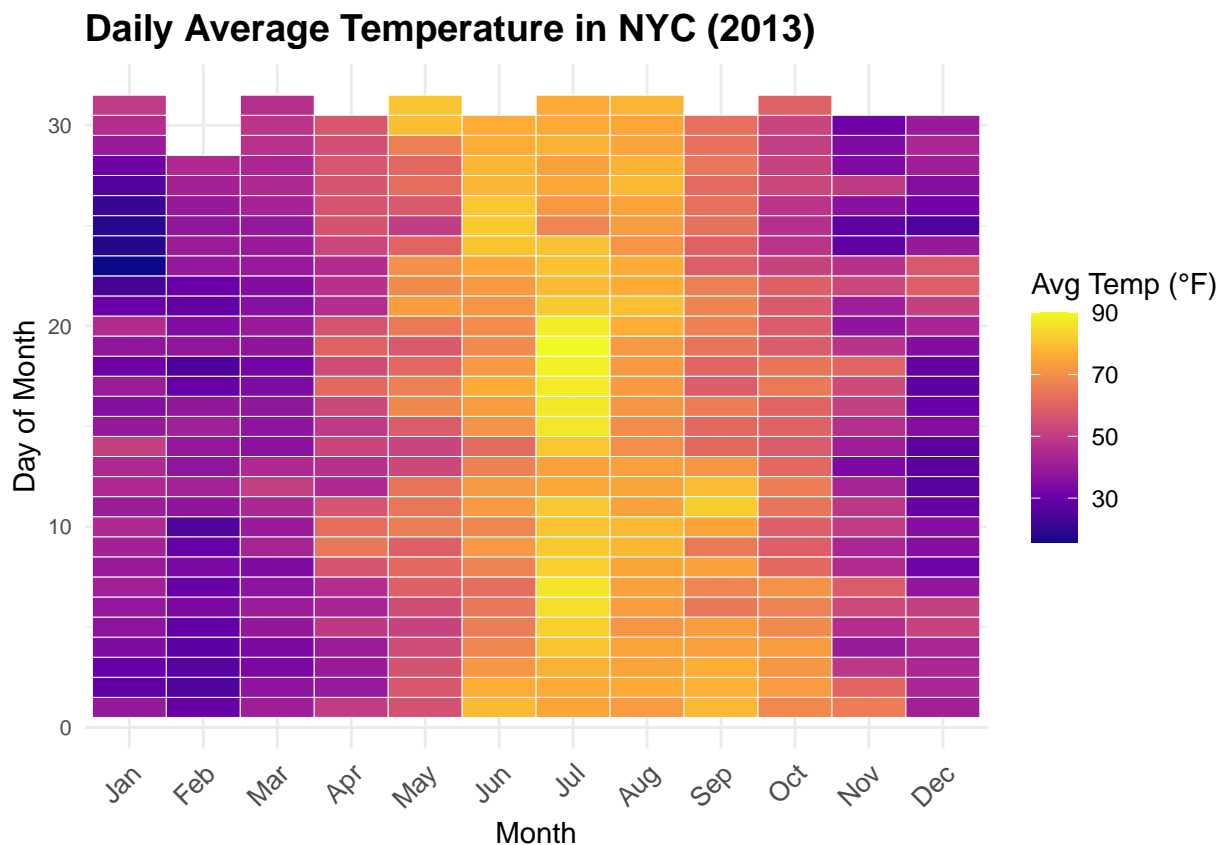
3. Weather Dataset

Using the weather dataset from the `nycflights13` package, create a plot to visualize the average daily temperature at NYC airports over the months of 2013.

```
library(nycflights13)
# Ensure this package is installed: install.packages("nycflights13")
head(weather)

## # A tibble: 6 x 15
##   origin year month   day hour temp dewp humid wind_dir wind_speed wind_gust
##   <chr>   <int> <int> <int> <int> <dbl> <dbl> <dbl>   <dbl>    <dbl>    <dbl>
## 1 EWR     2013     1     1     1  39.0  26.1  59.4     270     10.4      NA
## 2 EWR     2013     1     1     2  39.0  27.0  61.6     250      8.06     NA
## 3 EWR     2013     1     1     3  39.0  28.0  64.4     240     11.5     NA
## 4 EWR     2013     1     1     4  39.9  28.0  62.2     250     12.7     NA
## 5 EWR     2013     1     1     5  39.0  28.0  64.4     260     12.7     NA
## 6 EWR     2013     1     1     6  37.9  28.0  67.2     240     11.5     NA
## # i 4 more variables: precip <dbl>, pressure <dbl>, visib <dbl>,
## #   time_hour <dtm>
```

Here is a sample:



4. Diamonds Dataset

Diamonds is a dataset containing the prices and other attributes of almost 54,000 diamonds. Using the diamonds dataset, create a plot to visualize the distribution of price (price) across cut (cut) and clarity (clarity).

```
# Load the Data
data("diamonds")
head(diamonds)
```

```
## # A tibble: 6 x 10
```

##	carat	cut	color	clarity	depth	table	price	x	y	z
##	<dbl>	<ord>	<ord>	<ord>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>
## 1	0.23	Ideal	E	SI2	61.5	55	326	3.95	3.98	2.43
## 2	0.21	Premium	E	SI1	59.8	61	326	3.89	3.84	2.31
## 3	0.23	Good	E	VS1	56.9	65	327	4.05	4.07	2.31
## 4	0.29	Premium	I	VS2	62.4	58	334	4.2	4.23	2.63
## 5	0.31	Good	J	SI2	63.3	58	335	4.34	4.35	2.75
## 6	0.24	Very Good	J	VVS2	62.8	57	336	3.94	3.96	2.48

Here are a few samples:

Distribution of Diamond Prices by Cut and Clarity

