

IST687 – Viz HW: air quality Analysis

Step 1: Load the data

We will use the airquality data set, which you should already have as part of your R installation.

Step 2: Clean the data

After you load the data, there will be some NAs in the data. You need to figure out what to do about those nasty NAs.

Step 3: Understand the data distribution

Create the following visualizations using ggplot:

- Histograms for each of the variables
- Boxplot for Ozone
- Boxplot for wind values (round the wind to get a good number of “buckets”)

Step 3: Explore how the data changes over time

First, make sure to create appropriate dates (this data was from 1973). Then create line charts for ozone, temp, wind and solar.R (one line chart for each, and then one chart with 4 lines, each having a different color). Create these visualizations using ggplot.

Note that for the chart with 4 lines, you need to think about how to effectively use the y-axis.

Step 4: Look at all the data via a Heatmap

Create a heatmap, with each day along the x-axis and ozone, temp, wind and solar.r along the y-axis, and days as rows along the y-axis. Great the heatmap using geom_tile (this defines the ggplot geometry to be ‘tiles’ as opposed to ‘lines’ and the other geometry we have previously used).

Note that you need to figure out how to show the relative change equally across all the variables.

Step 5: Look at all the data via a scatter chart

Create a scatter chart (using ggplot geom_point), with the x-axis representing the wind, the y-axis representing the temperature, the size of each dot representing the ozone and the color representing the solar.R

Step 6: Final Analysis

- Do you see any patterns after exploring the data?
- What was the most useful visualization?