

Reading and Summarizing Data



10/10 points earned (100%)

Quiz passed!

[Back to Week 4](#)



1 / 1
points

1.

Use the readr package to read the **daily_SPEC_2014.csv.bz2** data file in to R. This file contains daily levels of fine particulate matter (PM2.5) chemical constituents across the United States. The data are measured at a network of federal, state, and local monitors and assembled by the EPA.

In this dataset, the "Sample.Value" column provides the level of the indicated chemical constituent and the "Parameter.Name" column provides the name of the chemical constituent. The combination of a "State.Code", a "County.Code", and a "Site.Num", uniquely identifies a monitoring site (the location of which is provided by the "Latitude" and "Longitude" columns).

For all of the questions below, you can ignore the missing values in the dataset, so when taking averages, just remove the missing values before taking the average (i.e. you can use `na.rm = TRUE` in the `mean()` function)

What is average Arithmetic.Mean for "Bromine PM2.5 LC" in the state of Wisconsin in this dataset?



1 / 1
points

2.

Calculate the average of each chemical constituent across all states, monitoring sites and all time points.

Which constituent Parameter.Name has the highest average level?



1 / 1
points

3.

Which monitoring site has the highest average level of "Sulfate PM2.5 LC" across all time?

Indicate the state code, county code, and site number.



1 / 1
points

4.

What is the absolute difference in the average levels of "EC PM2.5 LC TOR" between the states California and Arizona, across all time and all monitoring sites?



1 / 1
points

5.

What is the median level of "OC PM2.5 LC TOR" in the western United States, across all time? Define western as any monitoring location that has a Longitude LESS THAN -100.



1 / 1
points

6.

Use the readxl package to read the file **aqs_sites.xlsx** into R (you may need to install the package first). This file contains metadata about each of the monitoring sites in the EPA's monitoring system. In particular, the "Land Use" and "Location Setting" variables contain information about what kinds of areas the monitors are located in (i.e. "residential" vs. "forest").

How many monitoring sites are labelled as both RESIDENTIAL for "Land Use" and SUBURBAN for "Location Setting"?



1 / 1
points

7.

What is the median level of "EC PM2.5 LC TOR" amongst monitoring sites that are labelled as both "RESIDENTIAL" and "SUBURBAN" in the eastern U.S., where eastern is defined as Longitude greater than or equal to -100?



1 / 1
points

8.

Amongst monitoring sites that are labeled as COMMERCIAL for "Land Use", which month of the year has the highest average levels of "Sulfate PM2.5 LC"?



1 / 1
points

9.

Take a look at the data for the monitoring site identified by State Code 6, County Code 65, and Site Number 8001 (this monitor is in California). At this monitor, **for how many days is the sum of "Sulfate PM2.5 LC" and "Total Nitrate PM2.5 LC" greater than 10?**

For each of the chemical constituents, there will be some dates that have multiple `Arithmetic Mean` values at this monitoring site. When there are multiple values on a given date, take the average of the constituent values for that date.



1 / 1
points