

You must submit your answer using HTML. Submit your test as a single file <student ID>.html (e.g. 123456789.html). Send the file to Johan.Latulippe@viu.ca

- Dataset 1: [Tornadoes](#)
- Dataset 2: [Airline](#)

1. Data manipulation with Tornadoes. Use the Tornadoes data.

- (a) How many tornadoes occurred in WA (Washington) or MS (Mississippi)?
- (b) How many tornadoes occurred in WA after 2012?
- (c) Get all tornadoes that occurred in WA in 2012, 2013, and 2014.
- (d) Return the 'mo' (month), 'yr' (year), and 'f' (F-scale) for all tornadoes that occurred in HI (Hawaii).
- (e) Arrange the tornadoes by date and time. What state had the most recent tornado?
- (f) Get the count of tornadoes by months.

2. Data manipulation with Airline. (Use the Airline data)

- (a) Subset: Find the carrier (UniqueCarrier), the flight number (FlightNum), and origin (origin) of all the flights that:
  - i. Departed (DepTime) after 10pm and flew (Dest) to Nashville ('BNA').
  - ii. departed (DepTime) after 10pm, but Originated (Origin) from Nashville 'BNA' **or** landed (Dest) in Memphis 'MEM'. If you don't get a list of four flight, something is wrong.
  - iii. were delayed (ArrDelay) by more two hours. Note that a negative delay is a flight that arrived early.
  - iv. arrived (ArrDelay) more than two hours late, but didn't leave late (DepDelay).
- (b) Arrange: make use of minus sign in front of the function to reverse order() and the function head() to get the relevant part: (Use the Airline data)
  - i. Sort flights to find the 5 most delayed (DepDelay) flights
  - ii. Sort flights to find the 5 least delayed (DepDelay) flights
  - iii. Sort flights by destination (Dest) and break ties by **descending** arrival delay (ArrDelay)

**Bonus: Experts Only.**

1. Transform using the Airline data: create a new data frame with only the columns 'DepDelay', 'ArrDelay', 'Origin', 'Dest', 'AirTime', and 'Distance'. In addition keep only the observation for flights that were delayed (DepDelay) by more than 1 hour. (Use the Airline data)
  - (a) Add a new column reporting departure delays minus the mean departure delay to the new data frame
  - (b) Convert the departure and arrival delays columns from minutes into hours
  - (c) Add a column with the average flight speed (in mph)
  - (d) Report your new table using the function summary()