

# getweatherdata

Aviel Fradkine

2022-11-30

## weatherdata file part

Manipulate weather station data to get what we want

```
station_inventory <- read_csv("station_inventory.csv", skip = 3)

# filter for stations where there is data for first and last years 1991 - 2021
station_inventory_1991to2021 <- station_inventory %>%
  filter(`DLY First Year` < 1992 & `DLY Last Year` > 2020)

# filter by latitude lower than 55 (to exclude more northern territories)
station_inventory_1991to2021_filtered <- station_inventory_1991to2021 %>%
  filter(`Latitude (Decimal Degrees)` < 55)

station_lists_1991to2021_filtered <- station_inventory_1991to2021 %>%
  select(cols = c(`Name`, `Province`, `Station ID`, `Climate ID`))

colnames(station_lists_1991to2021_filtered) <- c('Name', 'Province', 'Station ID', 'Climate ID')

write_csv(station_inventory_1991to2021_filtered, "C:\\Users\\Aviel\\Desktop\\Econ468project\\dailystations1991to2021filtered.csv")
write_csv(station_lists_1991to2021_filtered, "C:\\Users\\Aviel\\Desktop\\Econ468project\\stationnames1991to2021filtered.csv")
```

Nov 9 2022

Exploring the stations

```
daily <- read_csv("dailystations1991to2021filtered.csv")

daily_alberta <- daily[daily$Province == "ALBERTA", ]
daily_manitoba <- daily[daily$Province == "MANITOBA", ]
daily_saskatchewan <- daily[daily$Province == "SASKATCHEWAN", ]
daily_ontario <- daily[daily$Province == "ONTARIO", ]
daily_quebec <- daily[daily$Province == "QUEBEC", ]

all_stations_filtered <- rbind(daily_alberta, daily_manitoba, daily_saskatchewan, daily_ontario, daily_quebec)

rm(list = setdiff(ls(), "all_stations_filtered")) # keeps only all stations filtered object
```

Combine all the stations into one frame

```
setwd("C:/Users/Aviel/Desktop/Econ468project/all_filtered_stations_1991to2021/")
a <- list.files("C:/Users/Aviel/Desktop/Econ468project/all_filtered_stations_1991to2021/")

length(a)
```

```
[1] 214
```

```
# read in all the station data
for (i in 1:length(a)){
  setwd("C:/Users/Aviel/Desktop/Econ468project/all_filtered_stations_1991to2021/")
  assign(print(a[i]), read_csv(a[i]))
}
```

```
[1] "amherstburg_4607.csv"
[1] "angers_5574.csv"
[1] "armstrong_(aut)_3987.csv"
[1] "arthabaska_5310.csv"
[1] "arundel_5575.csv"
[1] "athabasca_1_2459.csv"
[1] "bagotville_a_5889.csv"
[1] "baldur_3463.csv"
[1] "barrage_temiscamingue_5977.csv"
[1] "barwick_3932.csv"
[1] "beaver_3775.csv"
[1] "beechy_3071.csv"
[1] "belleville_4859.csv"
[1] "berens_river_cs_3721.csv"
[1] "bickleigh_3210.csv"
[1] "bonner_lake_4140.csv"
[1] "bonsecours_5322.csv"
[1] "breton_1825.csv"
[1] "broadview_2855.csv"
[1] "brockville_pcc_4236.csv"
[1] "bromptonville_5327.csv"
[1] "brooks_2180.csv"
[1] "brule_black_cat_2485.csv"
[1] "buffalo_pound_lake_2859.csv"
[1] "butte_st_pierre_3212.csv"
[1] "camrose_1839.csv"
[1] "cap_des_rosiers_5779.csv"
[1] "cape_whittle_5672.csv"
[1] "causapscal_5785.csv"
[1] "centreville_4898.csv"
[1] "chalk_river_aecl_4243.csv"
[1] "charteris_5584.csv"
[1] "chateau_richer_5674.csv"
[1] "chelsea_5585.csv"
[1] "cheneville_5586.csv"
[1] "chute-des-passes_5903.csv"
[1] "claresholm_2224.csv"
[1] "coaticook_5339.csv"
[1] "cobourg_stp_4905.csv"
[1] "cold_lake_a_2832.csv"
```

[1] "coldwater\_warminster\_4432.csv"  
[1] "cornwall\_4255.csv"  
[1] "coronach\_spc\_3172.csv"  
[1] "coteau\_du\_lac\_5217.csv"  
[1] "cowan\_3766.csv"  
[1] "danville\_5345.csv"  
[1] "deschambault\_5220.csv"  
[1] "drummond\_centre\_4268.csv"  
[1] "drummondville\_5348.csv"  
[1] "duval\_2890.csv"  
[1] "edmonton\_stony\_plain\_1870.csv"  
[1] "elbow\_2\_ne\_3091.csv"  
[1] "elbow\_cs\_3088.csv"  
[1] "elk\_island\_nat\_park\_1873.csv"  
[1] "elkhorn\_2\_east\_3460.csv"  
[1] "entwistle\_2513.csv"  
[1] "fabyan\_1881.csv"  
[1] "farnham\_5358.csv"  
[1] "fergus\_moe\_4761.csv"  
[1] "fergus\_shand\_dam\_4760.csv"  
[1] "fleury\_5361.csv"  
[1] "flin\_flon\_3857.csv"  
[1] "fort\_erie\_4635.csv"  
[1] "fort\_saskatchewan\_1886.csv"  
[1] "fortierville\_5362.csv"  
[1] "georgetown\_wctp\_4923.csv"  
[1] "georgeville\_5368.csv"  
[1] "granby\_5369.csv"  
[1] "gretna\_(aut)\_3605.csv"  
[1] "haliburton\_3\_5170.csv"  
[1] "hartington\_ihd\_4287.csv"  
[1] "hemmingford\_four\_winds\_5373.csv"  
[1] "herouxville\_5225.csv"  
[1] "hoey\_3300.csv"  
[1] "holland\_3609.csv"  
[1] "indian\_bay\_3746.csv"  
[1] "indian\_head\_cda\_2925.csv"  
[1] "jonquiere\_5911.csv"  
[1] "kananaskis\_2402.csv"  
[1] "kelliher\_2930.csv"  
[1] "kincardine\_4575.csv"  
[1] "kingsville\_moe\_4647.csv"  
[1] "kipling\_2933.csv"  
[1] "l'assomption\_5237.csv"  
[1] "la\_macaza\_5599.csv"  
[1] "lac\_aux\_sables\_5203.csv"  
[1] "lac\_berry\_6005.csv"  
[1] "lac\_humqui\_5803.csv"  
[1] "lac\_megantic\_2\_5385.csv"  
[1] "lachute\_5594.csv"  
[1] "laprairie\_5389.csv"  
[1] "last\_mountain\_cs\_2942.csv"  
[1] "laurierville\_5392.csv"  
[1] "lauzon\_5393.csv"

[1] "lennoxville\_5397.csv"  
 [1] "leroy\_2947.csv"  
 [1] "les\_buissons\_5701.csv"  
 [1] "les\_cedres\_5239.csv"  
 [1] "lethbridge\_cda\_2265.csv"  
 [1] "lingwick\_5398.csv"  
 [1] "lipton\_2\_2951.csv"  
 [1] "lloydminster\_a\_1920.csv"  
 [1] "louiseville\_5243.csv"  
 [1] "luskville\_5604.csv"  
 [1] "lyndhurst\_shawmere\_4308.csv"  
 [1] "mafeking\_3820.csv"  
 [1] "magog\_5401.csv"  
 [1] "mankota\_3186.csv"  
 [1] "marieville\_5406.csv"  
 [1] "marquette\_3619.csv"  
 [1] "massey\_4121.csv"  
 [1] "mccreary\_3853.csv"  
 [1] "monitor\_1935.csv"  
 [1] "mont\_brun\_5988.csv"  
 [1] "moosonee\_ua\_4168.csv"  
 [1] "muenster\_2973.csv"  
 [1] "new\_glasgow\_4656.csv"  
 [1] "notre\_dame de la paix\_5619.csv"  
 [1] "oakbank\_3641.csv"  
 [1] "oka\_5248.csv"  
 [1] "okotoks\_2173.csv"  
 [1] "ormstown\_5429.csv"  
 [1] "oshawa\_wpcp\_4996.csv"  
 [1] "ottawa\_cda\_4333.csv"  
 [1] "outlook\_pfra\_3318.csv"  
 [1] "oxbow\_2981.csv"  
 [1] "parent\_5966.csv"  
 [1] "parkerview\_2984.csv"  
 [1] "pickle\_lake (aut)\_3920.csv"  
 [1] "picture\_butte west\_2174.csv"  
 [1] "pierreville\_5432.csv"  
 [1] "pilot\_mound (aut)\_3649.csv"  
 [1] "port\_colborne\_4671.csv"  
 [1] "portage\_la prairie cda\_3519.csv"  
 [1] "qu'appelle\_1\_2992.csv"  
 [1] "queenstown\_2295.csv"  
 [1] "rawson\_lake\_3966.csv"  
 [1] "rennie\_3755.csv"  
 [1] "richmond\_5440.csv"  
 [1] "ridgeville\_4680.csv"  
 [1] "rigaud\_5252.csv"  
 [1] "rimouski\_5836.csv"  
 [1] "rivers\_pettapiece\_3457.csv"  
 [1] "riviere\_au tonnerre\_5650.csv"  
 [1] "riviere\_des prairies\_5441.csv"  
 [1] "riviere\_verte ouest\_5254.csv"  
 [1] "roblin\_friesen 3 northwest\_6879.csv"  
 [1] "rock\_point\_3142.csv"

[1] "rockglen\_(aut)\_3193.csv"  
[1] "roseville\_4816.csv"  
[1] "rossburn\_4\_north\_3458.csv"  
[1] "sabrevois\_5444.csv"  
[1] "saint\_michel\_5861.csv"  
[1] "sawyerville\_nord\_5526.csv"  
[1] "schanzenfeld\_6889.csv"  
[1] "scott\_5527.csv"  
[1] "scott\_cda\_3259.csv"  
[1] "shanty\_bay\_4509.csv"  
[1] "simonette\_2793.csv"  
[1] "sonningdale\_3261.csv"  
[1] "sorel\_5532.csv"  
[1] "st.\_albert\_4377.csv"  
[1] "st\_alban\_5255.csv"  
[1] "st\_alexis des monts\_5256.csv"  
[1] "st\_ambroise\_5936.csv"  
[1] "st\_anicet\_5448.csv"  
[1] "st\_bernard de lacolle\_5452.csv"  
[1] "st\_camille wolfe\_5300.csv"  
[1] "st\_camille\_5846.csv"  
[1] "st\_charles de mandeville\_5262.csv"  
[1] "st\_come de liniere\_5457.csv"  
[1] "st\_come\_5268.csv"  
[1] "st\_ephrem\_5462.csv"  
[1] "st\_georges\_5485.csv"  
[1] "st\_hippolyte\_5630.csv"  
[1] "st\_hyacinthe\_2\_5492.csv"  
[1] "st\_jean de cherbourg\_5857.csv"  
[1] "st\_jerome\_5631.csv"  
[1] "st\_labre\_3667.csv"  
[1] "st\_ludger\_5500.csv"  
[1] "st\_malo d auckland\_5502.csv"  
[1] "st\_michel des saints\_5969.csv"  
[1] "st\_narcisse\_5281.csv"  
[1] "st\_nazaire\_5506.csv"  
[1] "st\_pierre de broughton\_5509.csv"  
[1] "st\_prime\_5940.csv"  
[1] "st\_prosper\_5510.csv"  
[1] "st\_severin\_5516.csv"  
[1] "st\_thomas wpcp\_4689.csv"  
[1] "ste\_anne de la perade\_5257.csv"  
[1] "ste\_anne du lac\_5627.csv"  
[1] "ste\_beatrix\_5259.csv"  
[1] "ste\_christine\_5266.csv"  
[1] "ste\_lucie\_5858.csv"  
[1] "ste\_martine\_5503.csv"  
[1] "stewart\_valley\_3151.csv"  
[1] "stony\_mountain\_3678.csv"  
[1] "sutton\_5538.csv"  
[1] "swift\_current\_cda\_3157.csv"  
[1] "taber\_2315.csv"  
[1] "thetford\_mines\_5542.csv"  
[1] "tillsonburg\_wwtp\_4699.csv"

```
[1] "trenton_a_5126.csv"
[1] "udora_4525.csv"
[1] "upsala_(aut)_4057.csv"
[1] "val-d'or_a_6081.csv"
[1] "vegreville_1977.csv"
[1] "vercheres_5564.csv"
[1] "wasagaming_3562.csv"
[1] "welcome_island_(aut)_4061.csv"
[1] "windsor_riverside_4715.csv"
[1] "woodstock_4835.csv"
[1] "wright_5642.csv"
[1] "wroxeter_4603.csv"
```

```
all_stations_data <- get(a[1])
```

```
for (i in 2:length(a)){
  all_stations_data <- rbind(all_stations_data, get(a[i]))
  print(i)
}
```

```
[1] 2
[1] 3
[1] 4
[1] 5
[1] 6
[1] 7
[1] 8
[1] 9
[1] 10
[1] 11
[1] 12
[1] 13
[1] 14
[1] 15
[1] 16
[1] 17
[1] 18
[1] 19
[1] 20
[1] 21
[1] 22
[1] 23
[1] 24
[1] 25
[1] 26
[1] 27
[1] 28
[1] 29
[1] 30
[1] 31
[1] 32
[1] 33
[1] 34
[1] 35
```

[1] 36  
[1] 37  
[1] 38  
[1] 39  
[1] 40  
[1] 41  
[1] 42  
[1] 43  
[1] 44  
[1] 45  
[1] 46  
[1] 47  
[1] 48  
[1] 49  
[1] 50  
[1] 51  
[1] 52  
[1] 53  
[1] 54  
[1] 55  
[1] 56  
[1] 57  
[1] 58  
[1] 59  
[1] 60  
[1] 61  
[1] 62  
[1] 63  
[1] 64  
[1] 65  
[1] 66  
[1] 67  
[1] 68  
[1] 69  
[1] 70  
[1] 71  
[1] 72  
[1] 73  
[1] 74  
[1] 75  
[1] 76  
[1] 77  
[1] 78  
[1] 79  
[1] 80  
[1] 81  
[1] 82  
[1] 83  
[1] 84  
[1] 85  
[1] 86  
[1] 87  
[1] 88  
[1] 89

[1] 90  
[1] 91  
[1] 92  
[1] 93  
[1] 94  
[1] 95  
[1] 96  
[1] 97  
[1] 98  
[1] 99  
[1] 100  
[1] 101  
[1] 102  
[1] 103  
[1] 104  
[1] 105  
[1] 106  
[1] 107  
[1] 108  
[1] 109  
[1] 110  
[1] 111  
[1] 112  
[1] 113  
[1] 114  
[1] 115  
[1] 116  
[1] 117  
[1] 118  
[1] 119  
[1] 120  
[1] 121  
[1] 122  
[1] 123  
[1] 124  
[1] 125  
[1] 126  
[1] 127  
[1] 128  
[1] 129  
[1] 130  
[1] 131  
[1] 132  
[1] 133  
[1] 134  
[1] 135  
[1] 136  
[1] 137  
[1] 138  
[1] 139  
[1] 140  
[1] 141  
[1] 142  
[1] 143



[1] 144  
[1] 145  
[1] 146  
[1] 147  
[1] 148  
[1] 149  
[1] 150  
[1] 151  
[1] 152  
[1] 153  
[1] 154  
[1] 155  
[1] 156  
[1] 157  
[1] 158  
[1] 159  
[1] 160  
[1] 161  
[1] 162  
[1] 163  
[1] 164  
[1] 165  
[1] 166  
[1] 167  
[1] 168  
[1] 169  
[1] 170  
[1] 171  
[1] 172  
[1] 173  
[1] 174  
[1] 175  
[1] 176  
[1] 177  
[1] 178  
[1] 179  
[1] 180  
[1] 181  
[1] 182  
[1] 183  
[1] 184  
[1] 185  
[1] 186  
[1] 187  
[1] 188  
[1] 189  
[1] 190  
[1] 191  
[1] 192  
[1] 193  
[1] 194  
[1] 195  
[1] 196  
[1] 197

```
[1] 198
[1] 199
[1] 200
[1] 201
[1] 202
[1] 203
[1] 204
[1] 205
[1] 206
[1] 207
[1] 208
[1] 209
[1] 210
[1] 211
[1] 212
[1] 213
[1] 214
```

```
# write.csv(all_stations_data, "C:/Users/Aviel/Desktop/Econ468project/all_stations_data.csv")

rm(list = setdiff(ls(), "all_stations_copy")) # remove all the other objects
```

## Data exploring

```
all_stations_data <- read.csv("C:/Users/Aviel/Desktop/Econ468project/all_stations_data.csv")

all_stations_copy <- all_stations_data %>%
  select(cols = -c(`...1`)) # remove unnecessary column

# create station name vector
station_IDs <- unique(all_stations_copy$station_id)

filtered_stations <- read_csv("C:/Users/Aviel/Desktop/Econ468project/stationnames1991to2021filtered.csv")

# gives us 214 rows with each station being only the ones we use
filtered_stations <- filtered_stations %>%
  filter(`Station ID` %in% station_IDs)

# count number of NAs for each stations mean temp data
n_NAs_by_station <- all_stations_copy %>%
  group_by(`station_id`) %>%
  summarize(number_of_NAs = sum(is.na(`mean_temp`)))

# create a third column with percentage NA
# create a fourth column with province code
n_NAs_by_station <- n_NAs_by_station %>%
  mutate(percent_NAs = `number_of_NAs`/nrow(n_NAs_by_station)) %>%
  mutate(prov = case_when(
    (station_id %in% c(1824:2833)) ~ "AB",
    (station_id %in% c(2854:3319)) ~ "SK",
    (station_id %in% c(3462:3858)) ~ "MB",
    (station_id %in% c(3920:5171)) ~ "ON",
```

```

        (station_id %in% c(5216:7000))~ "QC"
    ))

# fix missing case with Manitoba
n_NAs_by_station[51:53, 4] <- c("MB")

# fix missing case with Quebec
n_NAs_by_station[114, 4] <- c("QC")

# sort by number of NA's
n_NAs_by_station <- n_NAs_by_station %>%
  group_by(`prov`) %>%
  arrange(`percent_NAs`, .by_group = TRUE)

stations_alberta <- n_NAs_by_station %>%
  filter(`prov` == c("AB"))
stations_manitoba <- n_NAs_by_station %>%
  filter(`prov` == c("MB"))
stations_sask <- n_NAs_by_station %>%
  filter(`prov` == c("SK"))
stations_ontario <- n_NAs_by_station %>%
  filter(`prov` == c("ON"))
stations_quebec <- n_NAs_by_station %>%
  filter(`prov` == c("QC"))

# ASSEMBLE SAMPLE
# 17 stations from alberta
sample_alberta <- stations_alberta %>%
  slice(c(1:17))
# 28 stations from saskatchewan
sample_saskatchewan <- stations_sask %>%
  slice(c(1:28))
# 8 stations from manitoba
sample_manitoba <- stations_manitoba %>%
  slice(c(1:8))
# 11 stations from ontario
sample_ontario <- stations_ontario %>%
  slice(c(1:11))
# 3 stations from quebec
sample_quebec <- stations_quebec %>%
  slice(c(1:3))

test_sample_station_NAs <- rbind(sample_alberta, sample_saskatchewan, sample_manitoba, sample_ontario, sample_quebec)

test_sample_station_IDs <- test_sample_station_NAs$station_id

## Take all the data for those stations

all_stations_sample <- all_stations_copy %>%
  filter(`station_id` %in% test_sample_station_IDs)

```

Now, using the data from the filtered set, make the temperature index

```

all_stations_sample_weather <- all_stations_sample %>%
  select(`station_name`, `station_id`, `lat`, `lon`, `elev`,
        `date`, `year`, `month`, `day`, `max_temp`, `mean_temp`, `min_temp`, `total_precip`)

# create a third column which has the month year combo as a date
all_stations_sample_weather$month_year <- format(as.Date(all_stations_sample_weather$date), "%Y-%m")

# order by date
all_stations_sample_weather <- all_stations_sample_weather %>%
  group_by(`date`, `month_year`) %>%
  arrange(date, .group_by = T)

# write.csv(all_stations_sample_weather, "C:/Users/Aviel/Desktop/Econ468project/sample_weather.csv")

```

[NOV 14] IMPORTANT: To resume work from the ordered data set for all the stations we are using use data in "sample\_weather.csv"

```

## create a weather index with mean temp by day for min, max, and mean temp
# each day and the average of mean temp across all stations
avg_mean_temp_by_day <- all_stations_sample_weather %>%
  group_by(`month_year`, `date`) %>%
  summarize(avg_mean_temp = mean(mean_temp, na.rm = TRUE))

# each day and the average of min temp across all stations
avg_min_temp_by_day <- all_stations_sample_weather %>%
  group_by(`month_year`, `date`) %>%
  summarize(avg_min_temp = mean(min_temp, na.rm = TRUE))

# each day and the average of max temp across all stations
avg_max_temp_by_day <- all_stations_sample_weather %>%
  group_by(`month_year`, `date`) %>%
  summarize(avg_max_temp = mean(max_temp, na.rm = TRUE))

# create a factor for the mean temp by day for month
avg_mean_temp_by_month <- avg_mean_temp_by_day %>%
  group_by(month_year) %>%
  summarize(month_mean_temp = mean(avg_mean_temp)) %>%
  arrange(month_year)

# write.csv(avg_mean_temp_by_month, "C:/Users/Aviel/Desktop/Econ468project/regression_inputs/monthly_mean_temp.csv")

```

Constructing the index for max temp

```

daily_max_30 <- all_stations_sample_weather %>%
  summarize(n_days_max_over_30 = sum(max_temp >= 30, na.rm = TRUE))

daily_max_25 <- all_stations_sample_weather %>%
  summarize(n_days_max_over_25 = sum(max_temp >= 25, na.rm = TRUE))

# summarize over months the number of stations in month per days which had over 30
monthly_max_30 <- daily_max_30 %>%
  group_by(`month_year`) %>%
  summarize(total_station_max_days_over_30 = sum(n_days_max_over_30))

# write.csv(monthly_max_30, "C:/Users/Aviel/Desktop/Econ468project/regression_inputs/monthly_max_30.csv")

```

```
# summarize over months the number of stations in month per days which had over 25
monthly_max_25 <- daily_max_25 %>%
  group_by(`month_year`) %>%
  summarize(total_station_max_days_over_25 = sum(n_days_max_over_25))
# write.csv(monthly_max_25, "C:/Users/Aviel/Desktop/Econ468project/regression_inputs/monthly_max_25.csv")
```

Now let's try to construct the index for mean

```
#data set
daily_mean_30 <- all_stations_sample_weather %>%
  summarize(n_days_over_30 = sum(mean_temp >= 30, na.rm = TRUE))

daily_mean_25 <- all_stations_sample_weather %>%
  summarize(n_days_over_25 = sum(mean_temp >= 25, na.rm = TRUE))

# summarize over months the number of stations in month per days which had over 30
monthly_mean_30 <- daily_mean_30 %>%
  group_by(`month_year`) %>%
  summarize(total_station_days_over_30 = sum(n_days_over_30))
# write.csv(monthly_mean_30, "C:/Users/Aviel/Desktop/Econ468project/regression_inputs/monthly_over_30.csv")

# summarize over months the number of stations in month per days which had over 25
monthly_mean_25 <- daily_mean_25 %>%
  group_by(`month_year`) %>%
  summarize(total_station_days_over_25 = sum(n_days_over_25))
# write.csv(monthly_mean_25, "C:/Users/Aviel/Desktop/Econ468project/regression_inputs/monthly_over_25.csv")
```

Precipitation data

```
precipitation_monthly <- all_stations_sample_weather %>%
  group_by(`month_year`) %>%
  summarize(monthly_precip = sum(total_precip, .by_group = T, na.rm = T)) %>%
  mutate(avg_precip = monthly_precip/length(unique(all_stations_sample_weather$station_name)))
# write.csv(precipitation_monthly, "C:/Users/Aviel/Desktop/Econ468project/regression_inputs/monthly_precip.csv")
```