

```
In [ ]: setwd("/home/leoKraushaar/Documents/School/Year 3/Semester 2/STAT 413/Project/protests/")
```

```
In [ ]: library(dplyr)
```

```
In [ ]: data <- read.csv("data/clean/canadianProtestData.csv")[, -1]
colnames(data)[colnames(data) == 'prov'] <- 'GEO'
```

```
In [ ]: dim(grouped_data)
```

299 · 3

```
In [ ]: retail <- read.csv("data/clean/retail.csv")
retail <- dplyr::rename(retail, ret_sales_tot=VALUE)

retail <- distinct(retail)
retail <- na.omit(retail)

retail_mean <- retail %>%
  group_by(REF_DATE, GEO) %>%
  summarise(retail = mean(ret_sales_tot, na.rm = TRUE))
```

`summarise()` has grouped output by 'REF\_DATE'. You can override using the  
`.groups` argument.

```
In [ ]: oil <- read.csv("data/clean/oil.csv")

oil <- dplyr::rename(oil, import_oil=VALUE)

oil <- distinct(oil)
oil <- na.omit(oil)

oil_mean <- oil %>%
  group_by(REF_DATE, GEO) %>%
  summarise(oil = mean(import_oil, na.rm = TRUE))

oil_mean
```

`summarise()` has grouped output by 'REF\_DATE'. You can override using the  
`.groups` argument.

A grouped\_df: 299 × 3

REF_DATE	GEO	oil
<chr>	<chr>	<dbl>
2022-01	Alberta	2442
2022-01	British Columbia	69664
2022-01	Manitoba	2893
2022-01	New Brunswick	1274
2022-01	Newfoundland and Labrador	132480
2022-01	Northwest Territories	0
2022-01	Nova Scotia	144972
2022-01	Nunavut	0
2022-01	Ontario	153320
2022-01	Prince Edward Island	0
2022-01	Quebec	94466
2022-01	Saskatchewan	1316
2022-01	Yukon	71
2022-02	Alberta	2311
2022-02	British Columbia	56639
2022-02	Manitoba	3877
2022-02	New Brunswick	33734
2022-02	Newfoundland and Labrador	81776
2022-02	Northwest Territories	0
2022-02	Nova Scotia	0
2022-02	Nunavut	0
2022-02	Ontario	114166
2022-02	Prince Edward Island	0
2022-02	Quebec	261073
2022-02	Saskatchewan	1616
2022-02	Yukon	0
2022-03	Alberta	4781
2022-03	British Columbia	64115
2022-03	Manitoba	6882
2022-03	New Brunswick	67036
⋮	⋮	⋮
2023-09	Prince Edward Island	0
2023-09	Quebec	327837
2023-09	Saskatchewan	1250
2023-09	Yukon	0
2023-10	Alberta	2004
2023-10	British Columbia	144566
2023-10	Manitoba	4114
2023-10	New Brunswick	1137
2023-10	Newfoundland and Labrador	0
2023-10	Northwest Territories	0
2023-10	Nova Scotia	63643
2023-10	Nunavut	0
2023-10	Ontario	110879
2023-10	Prince Edward Island	0
2023-10	Quebec	267734
2023-10	Saskatchewan	1628

REF_DATE	GEO	oil
<chr>	<chr>	<dbl>
2023-10	Yukon	0
2023-11	Alberta	2159
2023-11	British Columbia	75315
2023-11	Manitoba	3433
2023-11	New Brunswick	1239
2023-11	Newfoundland and Labrador	161520
2023-11	Northwest Territories	0
2023-11	Nova Scotia	20668
2023-11	Nunavut	76768
2023-11	Ontario	106134
2023-11	Prince Edward Island	0
2023-11	Quebec	311386
2023-11	Saskatchewan	1314
2023-11	Yukon	0

```
In [ ]: food <- read.csv("data/clean/food.csv")

food <- dplyr::rename(food, food_sales=VALUE)

food <- distinct(food)
food <- na.omit(food)

food_mean <- food %>%
  group_by(REF_DATE, GEO) %>%
  summarise(food = mean(food_sales, na.rm = TRUE))

food_mean
```

`summarise()` has grouped output by 'REF\_DATE'. You can override using the ``.groups`` argument.

A grouped\_df: 292 × 3

REF_DATE	GEO	food
<chr>	<chr>	<dbl>
2022-01	Alberta	759775
2022-01	British Columbia	1033526
2022-01	Manitoba	155613
2022-01	New Brunswick	91605
2022-01	Newfoundland and Labrador	64729
2022-01	Northwest Territories	4711
2022-01	Nova Scotia	129962
2022-01	Nunavut	1473
2022-01	Ontario	1916531
2022-01	Prince Edward Island	23022
2022-01	Quebec	884990
2022-01	Saskatchewan	157493
2022-01	Yukon	6125
2022-02	Alberta	843495
2022-02	British Columbia	1187563
2022-02	Manitoba	177460
2022-02	New Brunswick	110261
2022-02	Newfoundland and Labrador	73373
2022-02	Northwest Territories	5382
2022-02	Nova Scotia	153436
2022-02	Nunavut	2155
2022-02	Ontario	2423818
2022-02	Prince Edward Island	28763
2022-02	Quebec	1208015
2022-02	Saskatchewan	173912
2022-02	Yukon	6918
2022-03	Alberta	864944
2022-03	British Columbia	1202772
2022-03	Manitoba	193174
2022-03	New Brunswick	113005
⋮	⋮	⋮
2023-09	Prince Edward Island	35896
2023-09	Quebec	1528758
2023-09	Saskatchewan	196160
2023-09	Yukon	9221
2023-10	Alberta	1020938
2023-10	British Columbia	1360250
2023-10	Manitoba	214117
2023-10	New Brunswick	129902
2023-10	Newfoundland and Labrador	86490
2023-10	Northwest Territories	6509
2023-10	Nova Scotia	180974
2023-10	Nunavut	3715
2023-10	Ontario	3042934
2023-10	Prince Edward Island	35466
2023-10	Quebec	1534894
2023-10	Saskatchewan	198505

REF_DATE	GEO	food
<chr>	<chr>	<dbl>
2023-10	Yukon	9268
2023-11	Alberta	1019426
2023-11	British Columbia	1368515
2023-11	Manitoba	218308
2023-11	New Brunswick	130937
2023-11	Newfoundland and Labrador	86942
2023-11	Northwest Territories	6494
2023-11	Nova Scotia	185277
2023-11	Nunavut	4652
2023-11	Ontario	3084551
2023-11	Prince Edward Island	36342
2023-11	Quebec	1550373
2023-11	Saskatchewan	203970
2023-11	Yukon	9460

```
In [ ]: manufac <- read.csv("data/clean/manufac.csv")

manufac <- dplyr::rename(manufac, manufac_sales=VALUE)

manufac <- distinct(manufac)
manufac <- na.omit(manufac)

manufac_mean <- manufac %>%
  group_by(REF_DATE, GEO) %>%
  summarise(manufac = mean(manufac_sales, na.rm = TRUE))

manufac_mean
```

`summarise()` has grouped output by 'REF\_DATE'. You can override using the ``.groups`` argument.

A grouped_df: 324 × 3		
REF_DATE	GEO	manufac
<chr>	<chr>	<dbl>
2022-01	Alberta	8101486
2022-01	British Columbia	5570395
2022-01	Manitoba	1845320
2022-01	New Brunswick	2017001
2022-01	Newfoundland and Labrador	264436
2022-01	Northwest Territories	1435
2022-01	Northwest Territories including Nunavut	2066
2022-01	Nova Scotia	836638
2022-01	Nunavut	631
2022-01	Ontario	27714250
2022-01	Prince Edward Island	223492
2022-01	Quebec	16949941
2022-01	Saskatchewan	1776269
2022-01	Yukon	2767
2022-02	Alberta	8629092
2022-02	British Columbia	5832759
2022-02	Manitoba	2022267
2022-02	New Brunswick	2210765
2022-02	Newfoundland and Labrador	304927
2022-02	Northwest Territories	2029
2022-02	Northwest Territories including Nunavut	3965
2022-02	Nova Scotia	874999
2022-02	Nunavut	1936
2022-02	Ontario	30024531
2022-02	Prince Edward Island	223063
2022-02	Quebec	17442247
2022-02	Saskatchewan	1880361
2022-02	Yukon	2596
2022-03	Alberta	9081528
2022-03	British Columbia	6092015
⋮	⋮	⋮
2023-10	Nova Scotia	920235
2023-10	Ontario	31167250
2023-10	Prince Edward Island	284081
2023-10	Quebec	17796082
2023-10	Saskatchewan	2021620
2023-10	Yukon	3253
2023-11	Alberta	8740319
2023-11	British Columbia	5536415
2023-11	Manitoba	2218465
2023-11	New Brunswick	1810713
2023-11	Newfoundland and Labrador	274220
2023-11	Northwest Territories including Nunavut	3015
2023-11	Nova Scotia	893123
2023-11	Ontario	31833944
2023-11	Prince Edward Island	279229
2023-11	Quebec	17986302

REF_DATE	GEO	manufac
<chr>	<chr>	<dbl>
2023-11	Saskatchewan	2133737
2023-11	Yukon	4468
2023-12	Alberta	8870172
2023-12	British Columbia	5580420
2023-12	Manitoba	2103543
2023-12	New Brunswick	1978348
2023-12	Newfoundland and Labrador	222935
2023-12	Northwest Territories including Nunavut	3594
2023-12	Nova Scotia	893715
2023-12	Ontario	30954858
2023-12	Prince Edward Island	290573
2023-12	Quebec	18275178
2023-12	Saskatchewan	2039865
2023-12	Yukon	5121

```
In [ ]: power <- read.csv("data/clean/power.csv")
power <- power[power$Electric.power..components != "Total generation", ]
power <- power[, -c(3)]
power <- dplyr::rename(power, power_avail=VALUE)

power <- distinct(power)
power <- na.omit(power)

power_mean <- power %>%
  group_by(REF_DATE, GEO) %>%
  summarise(power = mean(power_avail, na.rm = TRUE))

power_mean
```

`summarise()` has grouped output by 'REF\_DATE'. You can override using the ``.groups`` argument.

A grouped_df: 299 × 3		
REF_DATE	GEO	power
<chr>	<chr>	<dbl>
2022-01	Alberta	6879472
2022-01	British Columbia	5996703
2022-01	Manitoba	2917343
2022-01	New Brunswick	1736016
2022-01	Newfoundland and Labrador	887826
2022-01	Northwest Territories	70307
2022-01	Nova Scotia	1203542
2022-01	Nunavut	18626
2022-01	Ontario	13360681
2022-01	Prince Edward Island	173892
2022-01	Quebec	25839325
2022-01	Saskatchewan	2361940
2022-01	Yukon	62834
2022-02	Alberta	5952095
2022-02	British Columbia	5675761
2022-02	Manitoba	2644092
2022-02	New Brunswick	1483608
2022-02	Newfoundland and Labrador	752737
2022-02	Northwest Territories	65084
2022-02	Nova Scotia	1060070
2022-02	Nunavut	17178
2022-02	Ontario	11683200
2022-02	Prince Edward Island	138145
2022-02	Quebec	21890451
2022-02	Saskatchewan	2032740
2022-02	Yukon	50853
2022-03	Alberta	6424649
2022-03	British Columbia	5867422
2022-03	Manitoba	2463401
2022-03	New Brunswick	1432127
⋮	⋮	⋮
2023-09	Prince Edward Island	117725
2023-09	Quebec	14563138
2023-09	Saskatchewan	1892052
2023-09	Yukon	37029
2023-10	Alberta	6576535
2023-10	British Columbia	5481719
2023-10	Manitoba	1883566
2023-10	New Brunswick	982231
2023-10	Newfoundland and Labrador	817346
2023-10	Northwest Territories	47285
2023-10	Nova Scotia	825066
2023-10	Nunavut	16165
2023-10	Ontario	11422274
2023-10	Prince Edward Island	118057
2023-10	Quebec	15244058
2023-10	Saskatchewan	1979795



REF_DATE	GEO	power
<chr>	<chr>	<dbl>
2023-10	Yukon	50892
2023-11	Alberta	6904386
2023-11	British Columbia	6211248
2023-11	Manitoba	2189819
2023-11	New Brunswick	1228522
2023-11	Newfoundland and Labrador	1050354
2023-11	Northwest Territories	56018
2023-11	Nova Scotia	968366
2023-11	Nunavut	17028
2023-11	Ontario	12093483
2023-11	Prince Edward Island	139348
2023-11	Quebec	18782951
2023-11	Saskatchewan	2199838
2023-11	Yukon	63017

```
In [ ]: date_as_string <- function(date) {
  month_map <- c("January", "February", "March", "April", "May", "June",
                "July", "August", "September", "October", "November", "December")
  names(month_map) <- 1:12

  year <- substr(date, start=1, stop=4)
  month <- substr(date, start=6, stop=nchar(date))

  if (substr(month, start=1, stop=1) == "0") {
    month <- substr(month, start=2, stop=nchar(month))
  }

  month_string <- month_map[as.integer(month)]

  return(c(year, month_string))
}

In [ ]: date_data <- as.data.frame(t(sapply(retail_mean$REF_DATE, FUN=date_as_string)))
retail_mean[, "year"] <- date_data$V1
retail_mean[, "month"] <- date_data$"1"

retail_mean$REF_DATE <- NULL

In [ ]: date_data <- as.data.frame(t(sapply(new_retail_mean$REF_DATE, FUN=date_as_string)))
new_retail_mean[, "year"] <- date_data$V1
new_retail_mean[, "month"] <- date_data$"1"

new_retail_mean$REF_DATE <- NULL

In [ ]: date_data <- as.data.frame(t(sapply(oil_mean$REF_DATE, FUN=date_as_string)))
oil_mean[, "year"] <- date_data$V1
oil_mean[, "month"] <- date_data$"1"

oil_mean$REF_DATE <- NULL

In [ ]: date_data <- as.data.frame(t(sapply(food_mean$REF_DATE, FUN=date_as_string)))
food_mean[, "year"] <- date_data$V1
food_mean[, "month"] <- date_data$"1"

food_mean$REF_DATE <- NULL

In [ ]: date_data <- as.data.frame(t(sapply(manufac_mean$REF_DATE, FUN=date_as_string)))
manufac_mean[, "year"] <- date_data$V1
manufac_mean[, "month"] <- date_data$"1"

manufac_mean$REF_DATE <- NULL

In [ ]: date_data <- as.data.frame(t(sapply(power_mean$REF_DATE, FUN=date_as_string)))
power_mean[, "year"] <- date_data$V1
power_mean[, "month"] <- date_data$"1"

power_mean$REF_DATE <- NULL

In [ ]: data <- merge(data, retail_mean, by = c("year", "month", "GEO"), all.x=TRUE)
data <- merge(data, oil_mean, by = c("year", "month", "GEO"), all.x=TRUE)
data <- merge(data, food_mean, by = c("year", "month", "GEO"), all.x=TRUE)
```

```
data <- merge(data, manufac_mean, by = c("year", "month", "GEO"), all.x=TRUE)
data <- merge(data, power_mean, by = c("year", "month", "GEO"), all.x=TRUE)

data
```

A data.frame: 299 × 10									
year	month	GEO	pop	protests	retail	oil	food	manufac	power
<int>	<chr>	<chr>	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2022	April	Alberta	4480956	17	7989056	3983	868863	9421575	6069621
2022	April	British Columbia	5310164	42	8959229	77433	1222442	5904432	5240902
2022	April	Manitoba	1405197	2	2083495	6290	194206	2000977	2168371
2022	April	New Brunswick	801778	5	1340707	1818	116742	2460804	1171958
2022	April	Newfoundland and Labrador	529249	2	920444	77160	78292	346772	686123
2022	April	Northwest Territories	44828	0	76390	0	5724	584	58889
2022	April	Nova Scotia	1014827	2	1689162	47821	164055	928882	899107
2022	April	Nunavut	40489	0	48635	0	1861	1856	16071
2022	April	Ontario	15046211	46	24616762	267687	2682207	31728284	10717875
2022	April	Prince Edward Island	165524	0	269014	0	30470	285508	129091
2022	April	Quebec	8627524	26	13896378	228362	1314059	18280632	17514950
2022	April	Saskatchewan	1173366	5	2048833	1285	178681	2120037	1869556
2022	April	Yukon	43454	2	88900	0	7819	3601	43778
2022	August	Alberta	4510891	6	8312320	4388	906253	8976474	6783590
2022	August	British Columbia	5356284	19	9144939	151630	1233372	5723411	4870866
2022	August	Manitoba	1413409	6	2231006	4489	198691	2156220	2054341
2022	August	New Brunswick	809568	4	1368487	2311	116593	2241491	997796
2022	August	Newfoundland and Labrador	531583	5	958844	150263	78940	264056	684606
2022	August	Northwest Territories	44685	0	75755	36714	5832	984	51834
2022	August	Nova Scotia	1025445	8	1714108	57149	166828	906456	847062
2022	August	Nunavut	40485	0	48163	0	2375	4295	14405
2022	August	Ontario	15145006	28	24320748	256796	2754474	30944157	12987686
2022	August	Prince Edward Island	167188	1	281978	0	33239	238437	115815
2022	August	Quebec	8672185	8	14351635	358902	1395697	17012905	15068395
2022	August	Saskatchewan	1178422	4	2110437	1606	181882	2175991	2063730
2022	August	Yukon	43905	0	92845	0	7493	3354	38095
2022	December	Alberta	4561350	4	8431294	2616	937555	8617749	7146188
2022	December	British Columbia	5403528	16	8960346	167172	1285432	5496946	6876391
2022	December	Manitoba	1423596	8	2275146	3608	208406	2056120	2876953
2022	December	New Brunswick	817766	7	1376248	1233	121964	2084289	1415122
:	:	:	:	:	:	:	:	:	:
2023	November	Prince Edward Island	175853	2	306554.00	0	36342	279229	139348
2023	November	Quebec	8948540	43	14861184.00	311386	1550373	17986302	18782951
2023	November	Saskatchewan	1218976	5	2118073.00	1314	203970	2133737	2199838
2023	November	Yukon	45148	4	97695.00	0	9460	4468	63017
2023	October	Alberta	4756408	17	8524706.00	2004	1020938	8664111	6576535
2023	October	British Columbia	5581127	31	9116046.00	144566	1360250	5398924	5481719
2023	October	Manitoba	1465440	14	2263706.00	4114	214117	2224327	1883566
2023	October	New Brunswick	842725	4	1470290.00	1137	129902	1918335	982231
2023	October	Newfoundland and Labrador	540418	7	953907.00	0	86490	259634	817346
2023	October	Northwest Territories	44760	0	83353.50	0	6509	NA	47285
2023	October	Nova Scotia	1066416	8	1774644.00	63643	180974	920235	825066
2023	October	Nunavut	40817	0	54694.50	0	3715	NA	16165
2023	October	Ontario	15801768	91	24940255.00	110879	3042934	31167250	11422274
2023	October	Prince Edward Island	175853	4	302505.00	0	35466	284081	118057
2023	October	Quebec	8948540	36	15090182.50	267734	1534894	17796082	15244058
2023	October	Saskatchewan	1218976	14	2180903.50	1628	198505	2021620	1979795

year	month	GEO	pop	protests	retail	oil	food	manufac	power
<int>	<chr>	<chr>	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2023	October	Yukon	45148	3	94912.00	0	9268	3253	50892
2023	September	Alberta	4695290	19	8548094.33	2956	1004084	9172187	6298895
2023	September	British Columbia	5519013	40	9073433.00	127585	1351233	5360919	4957719
2023	September	Manitoba	1454902	20	2255683.67	5158	212430	2244122	1744014
2023	September	New Brunswick	834691	12	1444919.67	24572	124672	2189054	905474
2023	September	Newfoundland and Labrador	538605	6	943546.33	123003	84021	274569	718676
2023	September	Northwest Territories	44972	0	68868.67	24660	5178	1411	48440
2023	September	Nova Scotia	1058694	13	1772434.00	42237	178627	889488	773087
2023	September	Nunavut	40673	0	54499.00	0	3058	2783	14630
2023	September	Ontario	15608369	73	24758223.67	149100	3009578	32607487	11407506
2023	September	Prince Edward Island	173787	4	304682.00	0	35896	279585	117725
2023	September	Quebec	8874683	22	14978844.33	327837	1528758	17810464	14563138
2023	September	Saskatchewan	1209107	7	2151378.33	1250	196160	2094269	1892052
2023	September	Yukon	44975	3	97186.67	0	9221	3892	37029

In [ ]: `dim(data)`

299 · 10

## New Retail

In [ ]: `new_retail <- read.csv("data/raw/more_retail.csv")[, c("GEO", "REF_DATE", "VALUE")]`

```
new_retail <- distinct(new_retail)
new_retail <- na.omit(new_retail)

new_retail_mean <- new_retail %>%
  group_by(REF_DATE, GEO) %>%
  summarise(retail = mean(VALUE, na.rm = TRUE))

head(new_retail_mean)
```

`summarise()` has grouped output by 'REF\_DATE'. You can override using the ``.groups`` argument.

A grouped\_df: 6 × 3

REF_DATE	GEO	retail
<chr>	<chr>	<dbl>
2017-01	Alberta	6726992
2017-01	British Columbia	7277591
2017-01	Manitoba	1749096
2017-01	New Brunswick	1049815
2017-01	Newfoundland and Labrador	800919
2017-01	Northwest Territories	65317

In [ ]: `date_data <- as.data.frame(t(sapply(new_retail_mean$REF_DATE, FUN=date_as_string)))`  
`new_retail_mean[, "year"] <- date_data$V1`  
`new_retail_mean[, "month"] <- date_data$"1"`  
  
`new_retail_mean$REF_DATE <- NULL`  
  
`head(new_retail_mean)`

A tibble: 6 × 4

GEO	retail	year	month
<chr>	<dbl>	<chr>	<chr>
Alberta	6726992	2017	January
British Columbia	7277591	2017	January
Manitoba	1749096	2017	January
New Brunswick	1049815	2017	January
Newfoundland and Labrador	800919	2017	January
Northwest Territories	65317	2017	January

```
In [ ]: write.csv(new_retail_mean, "data/clean/new_retail.csv")

In [ ]: # write.csv(data, "data/merged_data.csv")
```