capstone-1

junyou

2022-07-27

```
## read in files
spending_timeseries_t <- read_csv("~/Desktop/dpss/capstone-economy-fiscal/spending_timeseries_total.csv
spending_timeseries_p <- read_csv("~/Desktop/dpss/capstone-economy-fiscal/spending_timeseries_per.csv")
spending_maps <- read_csv("~/Desktop/dpss/capstone-economy-fiscal/spending_maps.csv")
spending_shift_in_baskets <- read_csv("~/Desktop/dpss/capstone-economy-fiscal/spending_shift_in_baskets
```

1. daily time series figure of spending by retail category

As we can see from the graph, from 2/1/2022 to 3/31/2022, spending in each categories share the same trend across the USA, with a soar of spending in weekend and relatively low spending in week-days. Among all categories, people spend most on general merchandise stores including warehouse clubs and supercenters, along with grocery stores, and people spend the least of their money on travel accommodation and clothing stores. After the war happened between Russia and Ukraine since 2/17/2022, people's consumption on general merchandise stores including warehouse clubs and supercenters and grocery stores has decreased, while people's consumption on gasoline has increased.

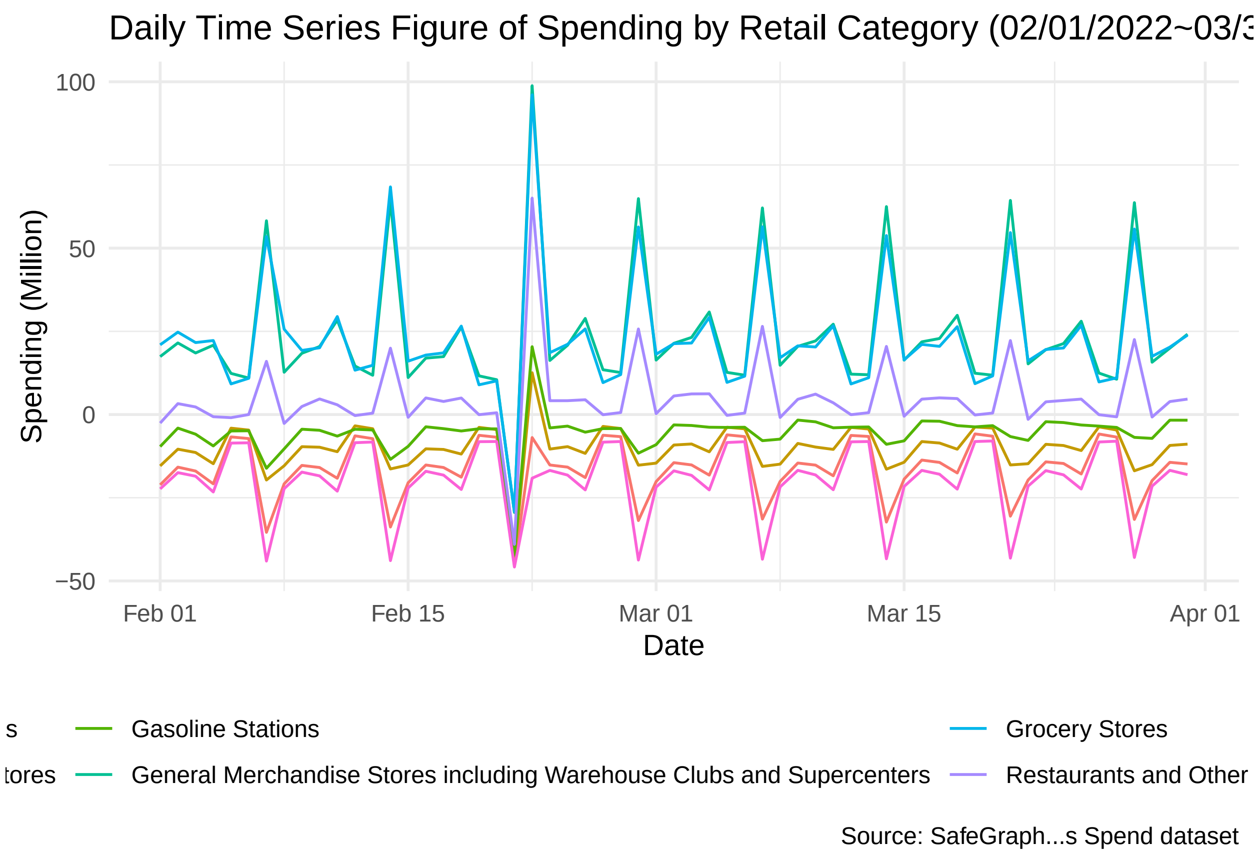
```
##daily time series figure of spending by retail category (spending, top_category)

spending_timeseries_t %>%
    mutate(date_smart = mdy(date_smart)) %>%
    ggplot(aes(x = date_smart, y = spending/1000000, color = `top_category`)) +
    geom_line() +
    labs(title = "Daily Time Series Figure of Spending by Retail Category (02/01/2022~03/31/2022)", x =
    y = "Spending (Million)", caption = "Source: SafeGraph's Spend dataset") +
    theme_minimal() +
    theme(legend.position = "bottom")
```

2. Deal with Cyclical Spending Patterns

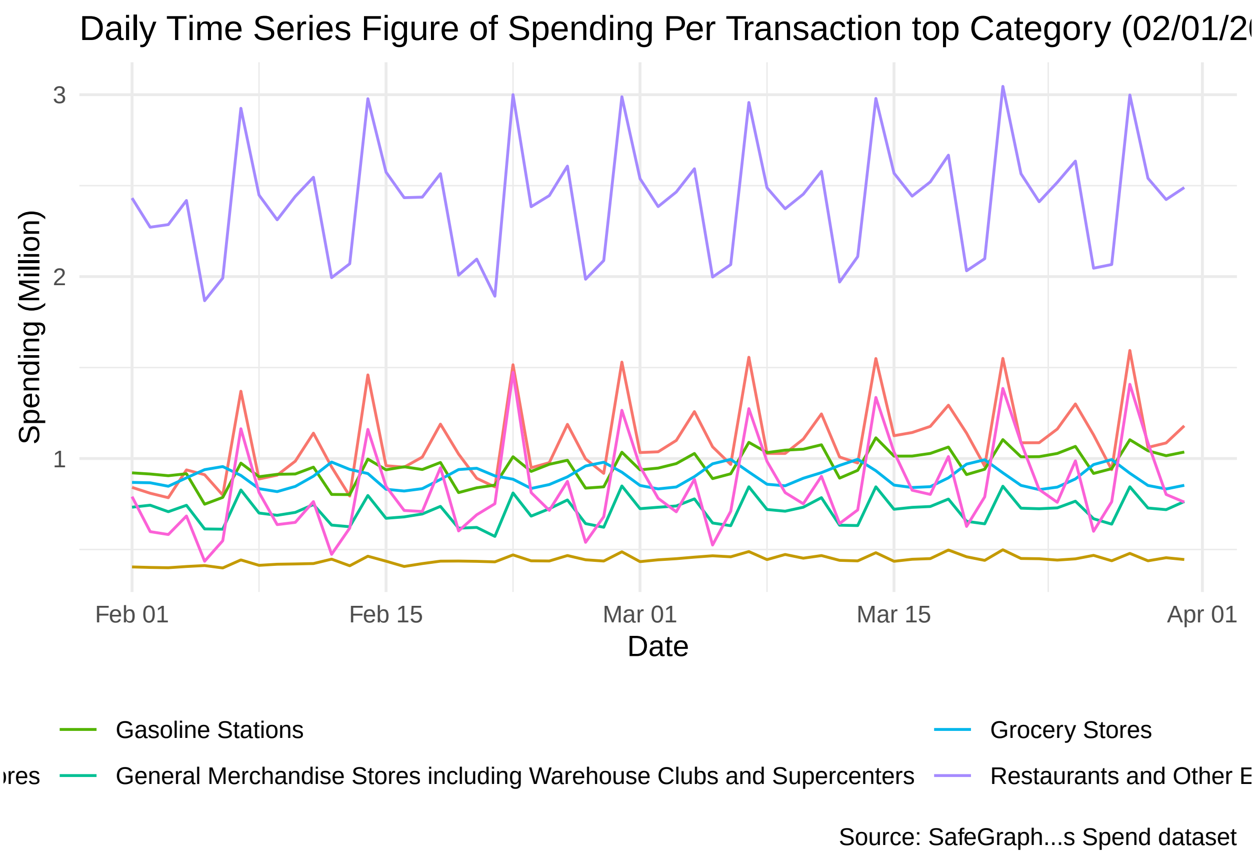
In the first graph, by demeaning the weekday in the same dataset as above, we removed the seasonality from the data, we can see that there is a surge right after February 21, 2022 when NATO member states announced sanction on Russia, among all categories, people costed significantly more on gasoline and grocery stores. In the second graph, by adopting the standardized spending by the number of transitions, which kept the median transaction size at this POI each day over the covered time period (the code book explanation can be found at here), surprisingly people spent significantly the most in restaurants and other eating places, among all categories, people's consumption on necessities in grocery stores didn't change much, while people's spending in restaurants, clothing stores, traveling accommodations steadily increased from February 1st to April 1st, 2022, which may be due to the gradual deregulation of the epidemic that might have boostered consumption. Around the time February 21, 2022 when NATO member states announced sanction on Russia, we can see that there's a decrease of spending on eating out, and a great increase on traveling accommodations, which might due to the increased oil prices and flight restrictions.

```
## PART 1

## demean approach
    ##demean wday
spending_timeseries_t %>%
    mutate(date_smart = mdy(date_smart)) %>%
    group_by(wday) %>%
    mutate(spending = demean(spending)) %>%
    ggplot(aes(x = date_smart, y = spending/1000000, color = `top_category`)) +
    geom_line() +
```

```
    labs(title = "Daily Time Series Figure of Spending by Retail Category (02/01/2022~03/31/2022)", x =
    y = "Spending (Million)", caption = "Source: SafeGraph's Spend dataset") +
    theme_minimal() +
    theme(legend.position = "bottom")
```




```
## part 2

spending_timeseries_p %>%
    mutate(date_smart = mdy(date_smart)) %>%
    ggplot(aes(x = date_smart, y = spending_ptpd/1000000, color = `top_category`)) +
    geom_line() +
    labs(title = "Daily Time Series Figure of Spending Per Transaction top Category (02/01/2022~03/31/20
    y = "Spending (Million)", caption = "Source: SafeGraph's Spend dataset") +
    theme_minimal() +
    theme(legend.position = "bottom")
```



3. Geographic mapping of the Post Invasion Change Before and After the Increase in Spending in Gasoline

As we can see from the graph, retail location mostly situated in big cities, especially in Chicago, where most transactions happen. The dot signifies the post invasion change before and after the increase in spending in gasoline in Illinois. Overall speaking, it seems that metropolitan areas are mostly densely impacted by the in crease in spending in Gasoline.

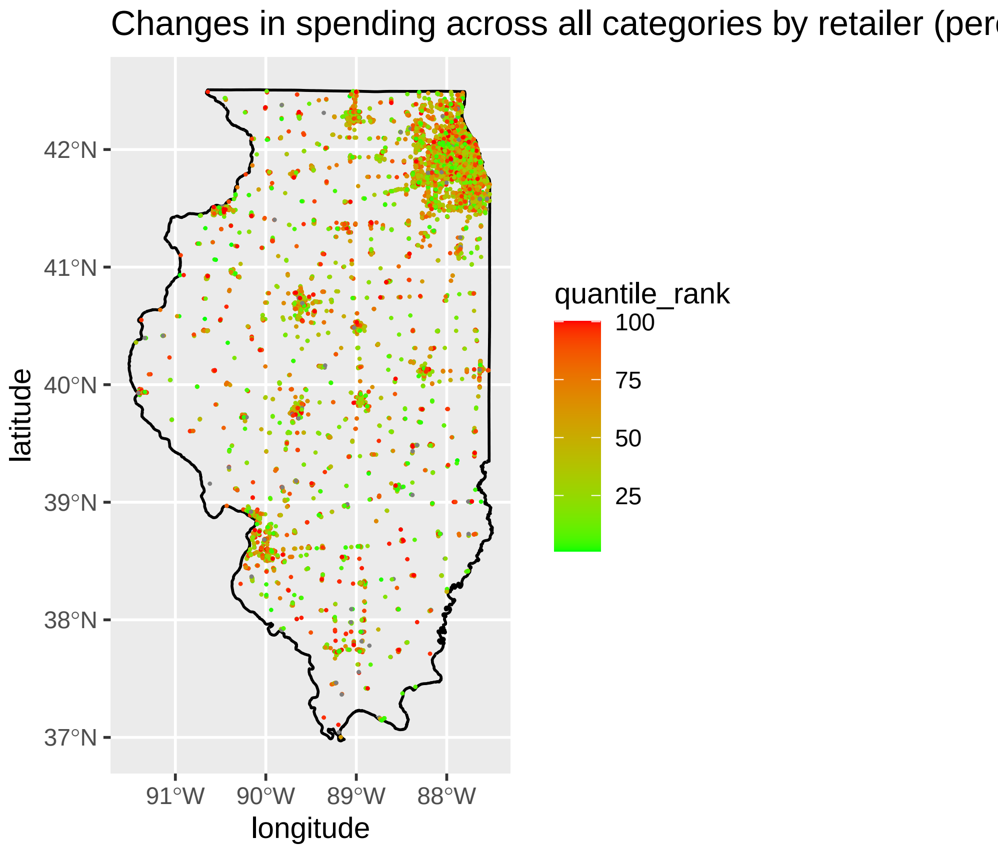
```
##shape file - illinois
##no boundaries
il_gdf_2 <- st_read("~/Desktop/IL_BNDY_State 2/IL_BNDY_State_Ln.shp")
```

```
## Reading layer ‘IL_BNDY_State_Ln’ from data source
##   ‘/Users/junyouchen/Desktop/IL_BNDY_State 2/IL_BNDY_State_Ln.shp’
##   using driver ‘ESRI Shapefile’
## Simple feature collection with 35 features and 1 field
## Geometry type: LINESTRING
## Dimension:     XY
## Bounding box:  xmin: -91.51352 ymin: 36.96997 xmax: -87.49521 ymax: 42.50835
## Geodetic CRS:  NAD83
```

```
##with boundaries
il_gdf <- st_read("~/Desktop/dpss/cb_2018_17_bg_500k/cb_2018_17_bg_500k.shp")
```

```
## Reading layer `cb_2018_17_bg_500k' from data source
##   `/Users/junyouchen/Desktop/dpss/cb_2018_17_bg_500k/cb_2018_17_bg_500k.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 9689 features and 10 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: -91.51308 ymin: 36.9703 xmax: -87.4952 ymax: 42.50848
## Geodetic CRS:  NAD83
```

```
il_gdf <- st_transform(il_gdf, 4326) %>%
    select(c(6,11))

# convert the latitude and longitude here into a geometry column
spending_m <- st_as_sf(
  spending_maps,
  coords = c("longitude", "latitude"),
  crs = 4326)

illinois_merged <- st_join(
  spending_m,  # points
  il_gdf,      # polygons
  join = st_within  # which polygon is the point WITHIN
  )

lonlat <- spending_maps %>%
    select(c(2:4))

## only keep GEO data in Illinois
illinois_merged <- illinois_merged %>%
    filter(!is.na(GEOID)) %>%
    left_join(lonlat, by = "placekey")

illinois_merged_9_quantile<-illinois_merged %>%
    mutate(quantile_rank = factor(ntile(illinois_merged$delta_STD, 9))) %>%
    mutate(quantile_rank = fct_rev(quantile_rank))

illinois_merged_100_quantile<-illinois_merged %>%
    mutate(quantile_rank = ntile(illinois_merged$delta_STD, 100))




ggplot() +
    geom_sf(data = il_gdf_2) +
    geom_point(
        data = illinois_merged_100_quantile, aes(x = longitude, y = latitude, color = quantile_rank), a
    scale_color_gradient(low = "green", high = "red") +
    labs(
        title = "Changes in spending across all categories by retailer (percentile of delta_STD) for Il
```



4. Regression to See How Changes in Spending on Gasoline Correlates with Changes in Spending on Other Goods

As we can see from the regression results, to study how the change in spending on gasoline by county correlates with changes in spending on all other goods and services in the same county, we first conducted a bivariate regression taking the post invasion change (5 business days) before and after the increase in spending in general goods (standardized) as the outcome and the post invasion change (5 business days) before and after the increase in spending in gasoline (standardized) as the regressor of interest, the result shows that changes in spending on gasoline is negatively correlated with changes in spending on all other goods and services, however, the correlation is statistically significant. We further conducted another linear regression model adding state fixed effects as well, the second regression still shows no statistical significance. Therefore, overall speaking, changes in spending on gasoline is not correlated with changes in spending on all other goods and services.

```
reg_nofe <- lm(delta_STD_gen ~ delta_STD_gas, data = spending_shift_in_baskets)
reg_statefe <- lm(delta_STD_gen ~ delta_STD_gas + as.factor(STATEFP), data = spending_shift_in_baskets)
```

```
reg_nofe %>%
    tidy() %>%
    kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	0.3535448	0.0639221	5.5308699	0.0000000
delta_STD_gas	-0.0058803	0.0271010	-0.2169775	0.8282421

```
reg_statefe %>%
    tidy() %>%
    kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	0.1265821	0.4056343	0.3120597	0.7550195
delta_STD_gas	-0.0055811	0.0274330	-0.2034435	0.8388039
as.factor(STATEFP)02	0.1417915	1.4147376	0.1002246	0.9201735
as.factor(STATEFP)04	0.1184364	0.9483231	0.1248903	0.9006198
as.factor(STATEFP)05	0.2687708	0.5775387	0.4653728	0.6417023
as.factor(STATEFP)06	0.5658856	0.6011003	0.9414164	0.3465764
as.factor(STATEFP)08	0.3245345	0.6071422	0.5345280	0.5930207
as.factor(STATEFP)09	0.1303542	1.2418603	0.1049669	0.9164100
as.factor(STATEFP)10	0.0446987	1.9591867	0.0228149	0.9817996
as.factor(STATEFP)11	0.6980780	3.3446308	0.2087160	0.8346858
as.factor(STATEFP)12	1.9467812	0.5735906	3.3940254	0.0006988
as.factor(STATEFP)13	0.1978119	0.4873372	0.4059036	0.6848459
as.factor(STATEFP)15	0.1615841	1.7087816	0.0945610	0.9246706
as.factor(STATEFP)16	0.0561220	0.6582777	0.0852559	0.9320643
as.factor(STATEFP)17	0.1735846	0.5396294	0.3216737	0.7477250
as.factor(STATEFP)18	0.0447348	0.5332041	0.0838981	0.9331437
as.factor(STATEFP)19	0.2049867	0.5758496	0.3559727	0.7218890
as.factor(STATEFP)20	0.2025229	0.6071811	0.3335462	0.7387481
as.factor(STATEFP)21	0.9867628	0.5078481	1.9430274	0.0521175
as.factor(STATEFP)22	0.2439402	0.5875337	0.4151936	0.6780334
as.factor(STATEFP)23	-0.0267185	0.9237750	-0.0289231	0.9769280
as.factor(STATEFP)24	0.1621477	0.7897890	0.2053051	0.8373494
as.factor(STATEFP)25	0.4256308	0.9755958	0.4362778	0.6626703
as.factor(STATEFP)26	0.0707923	0.5467468	0.1294791	0.8969883
as.factor(STATEFP)27	0.1496064	0.5674455	0.2636489	0.7920708
as.factor(STATEFP)28	0.0917104	0.5530267	0.1658337	0.8683004
as.factor(STATEFP)29	0.0997140	0.5332039	0.1870091	0.8516677
as.factor(STATEFP)30	0.1266390	0.6923944	0.1829001	0.8548902
as.factor(STATEFP)31	0.0995495	0.6356934	0.1565999	0.8755720
as.factor(STATEFP)32	0.2515560	0.9755883	0.2578505	0.7965421
as.factor(STATEFP)33	0.1188795	1.1254661	0.1056269	0.9158863
as.factor(STATEFP)34	-0.0030064	0.8302719	-0.0036210	0.9971111
as.factor(STATEFP)35	0.1576294	0.7379527	0.2136037	0.8308724
as.factor(STATEFP)36	0.2250791	0.5875513	0.3830800	0.7016909
as.factor(STATEFP)37	-0.0290714	0.5252062	-0.0553524	0.9558619
as.factor(STATEFP)38	-0.0641767	0.8304327	-0.0772810	0.9384058
as.factor(STATEFP)39	-0.0024824	0.5382947	-0.0046115	0.9963209
as.factor(STATEFP)40	0.0972309	0.5675411	0.1713195	0.8639855
as.factor(STATEFP)41	0.3326227	0.6990558	0.4758171	0.6342435
as.factor(STATEFP)42	-0.0198835	0.5736022	-0.0346643	0.9723500
as.factor(STATEFP)44	0.1512001	1.5391187	0.0982381	0.9217506
as.factor(STATEFP)45	0.2234450	0.6398675	0.3492052	0.7269627

term	estimate	std.error	statistic	p.value
as.factor(STATEFP)46	0.0129589	0.7897831	0.0164081	0.9869100
as.factor(STATEFP)47	0.2432476	0.5320699	0.4571724	0.6475842
as.factor(STATEFP)48	0.2424173	0.4623713	0.5242914	0.6001193
as.factor(STATEFP)49	0.2259109	0.7567854	0.2985139	0.7653341
as.factor(STATEFP)50	0.1044283	1.0061532	0.1037897	0.9173440
as.factor(STATEFP)51	0.0493735	0.4992861	0.0988881	0.9212345
as.factor(STATEFP)53	0.0960299	0.6860510	0.1399749	0.8886904
as.factor(STATEFP)54	0.0384146	0.6103007	0.0629438	0.9498160
as.factor(STATEFP)55	0.1291198	0.5654957	0.2283303	0.8194069
as.factor(STATEFP)56	0.3201057	0.8302693	0.3855445	0.6998647