Data Munging for *The Purchasing Plague: Modeling the Spread of Panic Buying*

This data munging was done for Anthony Ballestas's modeling research on panic purchasing during COVID-19 from February 2020 through July 2020.

PI: Anthony Ballestas (https://www.linkedin.com/in/anthony-ballestas/)
DS: Dominic Tarro (https://www.linkedin.com/in/dominic-tarro-1ab4b5152/)

Abstract

The Susceptible-Infected-Recovered (SIR) mathematical model is a well known system of ordinary differential equations typically used in the field of epidemiology to map the spread of infectious diseases. However, the SIR is a versatile model that has been applied in fields such as economics, sociology, and cybersecurity. In this paper, we modify and apply the SIR model to the spreading of panic buying throughout the state of Florida during the 2020 COVID-19 pandemic. After building the model, we tested its robustness by comparing it to Google movement data. The county by county movement data to grocery stores, as described by Google, denotes the number of visitors and length of stay on each day by tracking cell-phone locations. The mathematical model shows promise for modeling the spread of panic buying and the model further improves when considering stochastic variations. We are hoping to use our model to predict future waves of panic buying and extending our Florida-focused study to other states.

Group By Condition, Perform Count and Concatenate

Goal

- Count the Florida counties that have $grocery \geq 5$ before April 3, 2020 or $grocery \geq 0$ on or after April 3, 2020.
- Keep a list of those counties

```
In [1]:  import pandas as pd
         from datetime import datetime, timedelta

         df = pd.read_csv("Florida_grocery_movement.csv", index_col="Unnamed: 0")

         df
```

Out[1]:

	date	grocery	sub_region_2
284065	2/15/20	2.0	Alachua County
284066	2/16/20	-1.0	Alachua County
284067	2/17/20	-3.0	Alachua County
284068	2/18/20	0.0	Alachua County
284069	2/19/20	-2.0	Alachua County
...	...	...	...
293294	6/29/20	10.0	Washington County
293295	6/30/20	19.0	Washington County
293296	7/1/20	21.0	Washington County
293297	7/2/20	29.0	Washington County
293298	7/3/20	20.0	Washington County

7886 rows × 3 columns

```
In [2]: df["date"] = pd.to_datetime(df["date"], format="%m/%d/%y")
        df = df.sort_values(['date'])
        df.reset_index(inplace=True)
        df.drop(columns=['index'], inplace=True)

        # Date range
        start, end = df['date'].iloc[[0,-1]]
        print(start.date(), "->", end.date())
        print("Range of Days: ", (end-start).days + 1)

        # Pivot date
        pivot = datetime.strptime("04/03/20", "%m/%d/%y")
        print("Pivot Date: ", pivot.date())

        df
```

```
2020-02-15 -> 2020-07-03
Range of Days:  140
Pivot Date:  2020-04-03
```

Out[2]:

	date	grocery	sub_region_2
0	2020-02-15	2.0	Alachua County
1	2020-02-15	2.0	Hendry County
2	2020-02-15	0.0	Suwannee County
3	2020-02-15	-18.0	Hardee County
4	2020-02-15	-9.0	Hamilton County
...	...	...	...
7881	2020-07-03	2.0	Marion County
7882	2020-07-03	-12.0	Hardee County
7883	2020-07-03	4.0	Jackson County
7884	2020-07-03	18.0	Wakulla County
7885	2020-07-03	20.0	Washington County

7886 rows × 3 columns

```
In [3]: dfwork = df.copy()
        # Create dataframe of observations that meet one of two conditions
        dfwork = dfwork[((dfwork["date"] < pivot) & (dfwork["grocery"] >= 5)) |
                        ((dfwork["date"] >= pivot) & (dfwork["grocery"] >= 0))]
        dfwork
```

Out[3]:

	date	grocery	sub_region_2
5	2020-02-15	10.0	Gulf County
6	2020-02-15	5.0	Gilchrist County
8	2020-02-15	10.0	Franklin County
9	2020-02-15	6.0	Flagler County
10	2020-02-15	38.0	Taylor County
...	...	...	...
7879	2020-07-03	2.0	Brevard County
7881	2020-07-03	2.0	Marion County
7883	2020-07-03	4.0	Jackson County
7884	2020-07-03	18.0	Wakulla County
7885	2020-07-03	20.0	Washington County

2562 rows × 3 columns

In [4]:
```python
"""

On retained observations, group by the date and perform an operation on the county column:
    - count
    - concatenate

"""
dffinal = dfwork.groupby('date').sub_region_2.agg([('count', 'count'), ('counties', ','.join)])
dffinal
```

Out[4]:

date	count	counties
2020-02-15	16	Gulf County,Gilchrist County,Franklin County,F...
2020-02-16	12	Washington County,Pinellas County,Bradford Cou...
2020-02-17	9	Monroe County,Bradford County,Flagler County,G...
2020-02-18	14	Franklin County,Seminole County,St. Johns Coun...
2020-02-19	13	Gulf County,Jefferson County,Citrus County,Mon...
...	...	...
2020-06-29	16	Escambia County,Walton County,Clay County,Levy...
2020-06-30	16	Jackson County,Walton County,Putnam County,Was...
2020-07-01	20	Walton County,Putnam County,Taylor County,Clay...
2020-07-02	28	Walton County,Pasco County,Baker County,St. Jo...
2020-07-03	38	Pinellas County,Clay County,Pasco County,Volus...

137 rows × 2 columns

In [5]:
```python
dffinal.to_csv("./Florida_grocery_movement_county_condition.csv")
```