

W201 2020 Summer Project 2: US Diet Trends: 1970 to 2019

(Vegetables, Fruits, Meat and Dairy)

Team 5: Alan Zhang, Haoyu Zhang, Ratan Singh

Introduction

The ultimate question we want to answer for this project is whether the diet has changed over the past fifty years for an average American for food types such as fruit, vegetables, meat and dairy. We want to understand better if Americans now are eating the same way as our parents and grandparents.

We were able to answer our main question through the data points from USDA ERS site. We first collected the data points from these websites and then compared them against other sources, such as Statista to verify the accuracy. Then, we analyzed the data points by focusing on key categories, such as fresh and canning, and leading food types. As a team, we took the divide and conquer approach that each one of us took a section of food and performed analysis, and we brought those analysis together in the end.

Data Source & Structure

We relied on datasets from the United States Department of Agriculture - ERS sites, which included yearbooks for selected food types. We explored multiple datasets to collect a comprehensive list of data for all the food types we want to analyze. Most datasets contain the following data points:

- Per capita consumption/use & availability
- Utilized production and value of production
- Domestic use, production
- Average price indexes
- Import and Export values

Data Cleaning & Validation

The original datasets are organized according to categories as general, supply & utilization, U.S trade, fruit/vegetable/meat types, etc. With the help of lists of tables included in the dataset, it is possible to locate target data from hundreds of tables.

The main data cleaning work was to deal with missing or discontinued data in some categories. In most cases, we reassigned a numpy none value to the missing or discontinued data. As below, the missing data of imports/exports are treated as none value with assumption that data record of imports/exports were not available or hasn't started back then.

	Production	Imports	Beginning Stocks	Total Supply	Exports	Seed Use
Year						
1970	1,316.0	NaN	NaN	1,316.0	NaN	68.5
1971	1,149.4	NaN	158.6	1,308.0	NaN	68.0
1972	1,217.0	NaN	105.5	1,322.5	NaN	70.7
1973	1,215.6	NaN	141.3	1,356.9	NaN	80.5
1974	1,333.9	NaN	114.4	1,448.3	NaN	79.2
1975	1,289.1	NaN	219.7	1,508.8	NaN	87.6
1976	1,327.3	NaN	166.5	1,493.8	NaN	76.9
1977	1,188.5	NaN	129.1	1,317.6	NaN	91.6
1978	1,311.5	14.2	92.4	1,418.1	17.5	101.7
1979	1,337.0	13.1	131.2	1,481.3	19.4	92.5

In the situation that the plots are interrupted by individual missing data, the interpolate method is applied to acquire a connecting plots diagram.

In addition, there is some timeline difference between certain tables. For example, tables of supply and utilization of leaf lettuce and head lettuce are of different timeline.

	Production	Imports	Total Supply	Exports	Domestic Availability	Per Capita Availability	Current Dollars	Constant 2012 Dollars
Year								
1970	4,836.5	2.3	4,838.8	250.5	4,588.3	22.38	4.75	21.91
1971	4,936.7	4.5	4,941.2	292.8	4,648.4	22.38	6.31	27.70
1972	5,047.0	1.2	5,048.2	338.3	4,709.9	22.44	5.73	24.12
1973	5,243.5	1.9	5,245.4	345.5	4,899.9	23.12	7.40	29.53
1974	5,323.1	3.3	5,326.4	300.6	5,025.8	23.50	6.93	25.37
1975	5,410.8	2.2	5,413.0	329.6	5,083.4	23.54	6.71	22.48
1976	5,640.0	3.0	5,643.0	360.8	5,282.2	24.23	8.26	26.23
1977	6,043.2	3.8	6,047.0	359.5	5,687.5	25.82	6.94	20.75
1978	6,052.8	5.7	6,058.5	459.9	5,579.9	25.07	9.90	27.65
1979	6,143.9	13.0	6,156.9	480.6	5,648.5	25.10	9.20	23.73

	Production	Imports	Total Supply	Exports	Domestic Availability	Per Capita Availability	Current Dollars	Constant 2012 Dollars
Year								
1985	778.7	—	778.7	—	778.7	3.27	—	—
1986	571.2	—	571.2	—	571.2	2.37	—	—
1987	613.0	—	613.0	—	613.0	2.52	—	—
1988	784.2	—	784.2	—	784.2	3.20	—	—
1989	915.8	23.5	939.3	57.9	881.4	3.56	—	—
1990	1,061.6	12.1	1,073.7	130.6	943.1	3.77	15.60	24.50
1991	1,157.5	8.3	1,165.8	152.7	1,013.1	4.00	15.00	22.79
1992	1,388.7	5.9	1,394.6	195.0	1,199.6	4.67	20.66	30.69
1993	1,535.5	6.8	1,542.3	230.2	1,312.1	5.04	25.35	36.78
1994	1,710.0	8.9	1,718.9	223.1	1,495.8	5.68	23.92	33.98

To explore the total supply and utilization of lettuce, these two tables are combined by the operation `pandas.concat`. Meanwhile, the non-existing data for leaf lettuce from 1970 to 1989 are assigned to be 0. It is assumed that data of head lettuce is a close estimate of total lettuce when data of leaf lettuce is not available and no related comment is found from the original dataset. In this way, the acquired table for further analysis is shown below.

	Production	Imports	Exports	Domestic Availability	Production_b	Imports_b	Exports_b	Domestic Availability_b
1970	4836.5	2.3	250.5	4588.3	0.0	0.0	0.0	0.0
1971	4936.7	4.5	292.8	4648.4	0.0	0.0	0.0	0.0
1972	5047.0	1.2	338.3	4709.9	0.0	0.0	0.0	0.0
1973	5243.5	1.9	345.5	4899.9	0.0	0.0	0.0	0.0
1974	5323.1	3.3	300.6	5025.8	0.0	0.0	0.0	0.0
1975	5410.8	2.2	329.6	5083.4	0.0	0.0	0.0	0.0
1976	5640.0	3.0	360.8	5282.2	0.0	0.0	0.0	0.0
1977	6043.2	3.8	359.5	5687.5	0.0	0.0	0.0	0.0
1978	6052.8	5.7	459.9	5579.9	0.0	0.0	0.0	0.0
1979	6143.9	13.0	480.6	5648.5	0.0	0.0	0.0	0.0
1980	6336.3	15.1	488.5	5836.9	0.0	0.0	0.0	0.0
1981	6268.2	11.4	523.9	5728.6	0.0	0.0	0.0	0.0
1982	6294.9	14.6	499.3	5789.9	0.0	0.0	0.0	0.0
1983	5775.5	21.4	519.2	5258.6	0.0	0.0	0.0	0.0
1984	6397.6	32.6	524.1	5894.7	0.0	0.0	0.0	0.0
1985	6133.4	37.8	507.4	5644.9	778.7	0.0	0.0	778.7
1986	5829.0	20.8	553.6	5279.2	571.2	0.0	0.0	571.2
1987	6787.7	18.3	542.5	6242.2	613.0	0.0	0.0	613.0
1988	7050.5	37.4	431.3	6625.4	784.2	0.0	0.0	784.2
1989	7523.1	35.6	463.6	7095.1	915.8	23.5	57.9	881.4

We validated the data points against other sources such as Statista to make sure the data points make sense. We also studied the trends of the data as well as the descriptive metrics to make sure that the data is intact.

Vegetables

For vegetables, we explored two main datasets: the availability dataset and the imports/exports dataset. Availability is a function of domestic production, imports and exports and it represents the domestic supply of vegetables. We assume that we can use the availability dataset as a proxy for demand or consumption because we believe that the supply and demand should balance out in the long-term.

Veggie: Overall Availability

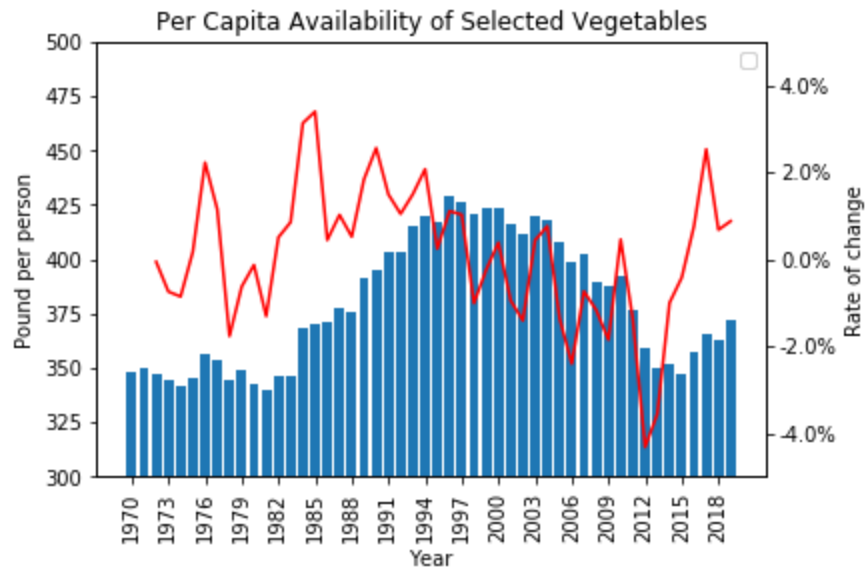


Figure: V1

We first looked into the overall availability of vegetables since 1970. As we can see in the figure above, the blue bar represents the overall availability of vegetables in the unit pound per capita and the red line represents the annual rate of change in the overall availability. In 1970, an average American had roughly 350 pound vegetables available. The per capita number peaked in 1996 at around 425 pounds per capita annually. Since then, the per capita availability has been decreasing until it rebounded in the last five years. Currently, the per capita vegetable availability is slightly higher than 50 years ago.

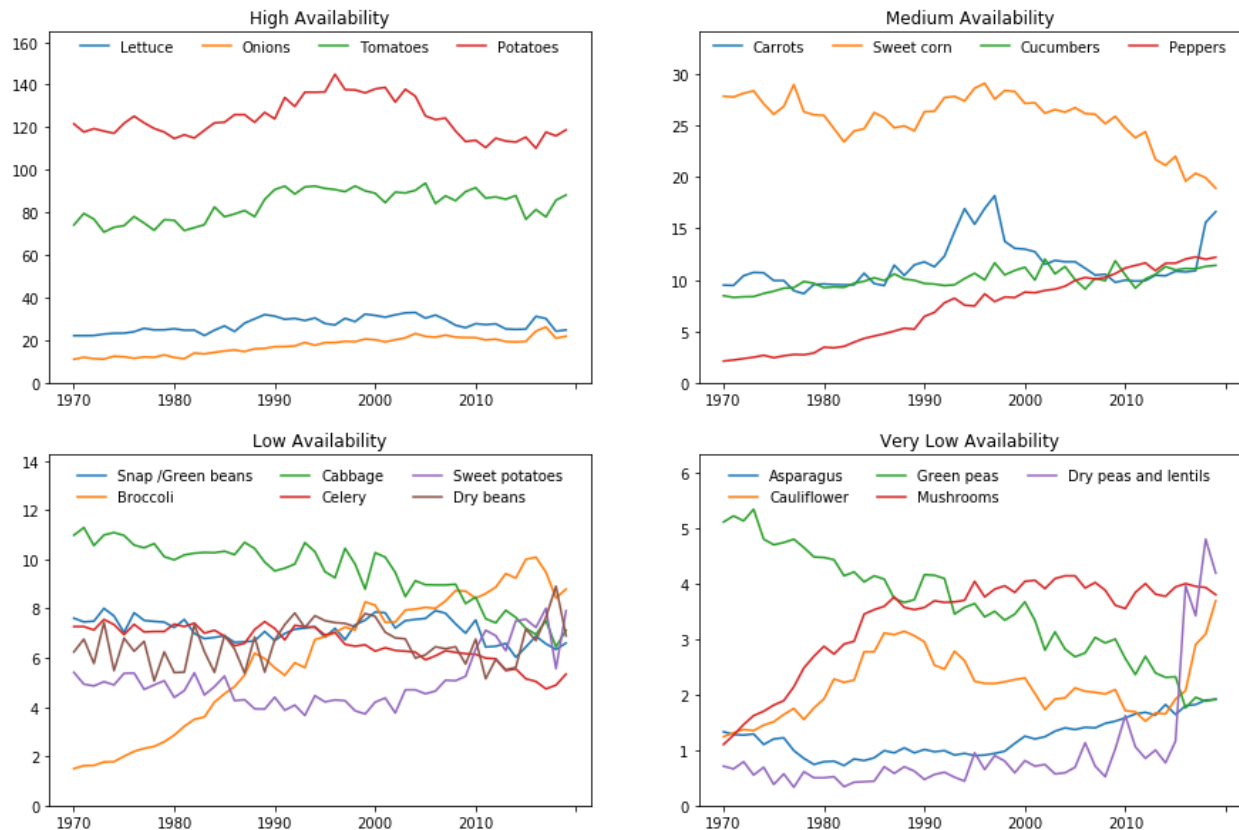


Figure: V2

Next, we analyzed the availability trends for each selected vegetable type. There are about 20 selected vegetable types and an all-inclusive chart is hard to analyze, so we broke the data points down into four different categories based on 2019 data:

1. High availability: > 20 pound per capita
2. Medium availability: 10 - 20 pound per capita
3. Low availability: 5 - 10 pound per capita
4. Very low availability: 0 - 5 pound per capita

From the two charts on top, we can identify the five most available vegetable types in 2019, which would be potatoes, tomatoes, lettuce, onions and sweet corn. The availability of sweet corn dropped in the last two decades but it remained in the top 5 group. For the four vegetables in the top right chart, they remained stable over time.

We could also identify some winners and losers in the charts above. We have a few vegetable types that have gained popularity over the years, such as broccoli, mushrooms, dry peas and lentils. We also have several vegetable types with declining availability such as sweet corn, cabbage and green peas.

Veggie: Process Methods

Another perspective to analyze the availability dataset is to focus our attention on the three main categories: Fresh, Canning and Freezing. The charts below broke out the three categories to demonstrate the overall trends for the three categories over the past fifty years. Freezing was quite stable, fluctuating around 5% of the total, while fresh gained momentum with a decline in the canning category. Over time, Americans grew the appetite for fresher vegetables and canning products took a hit.

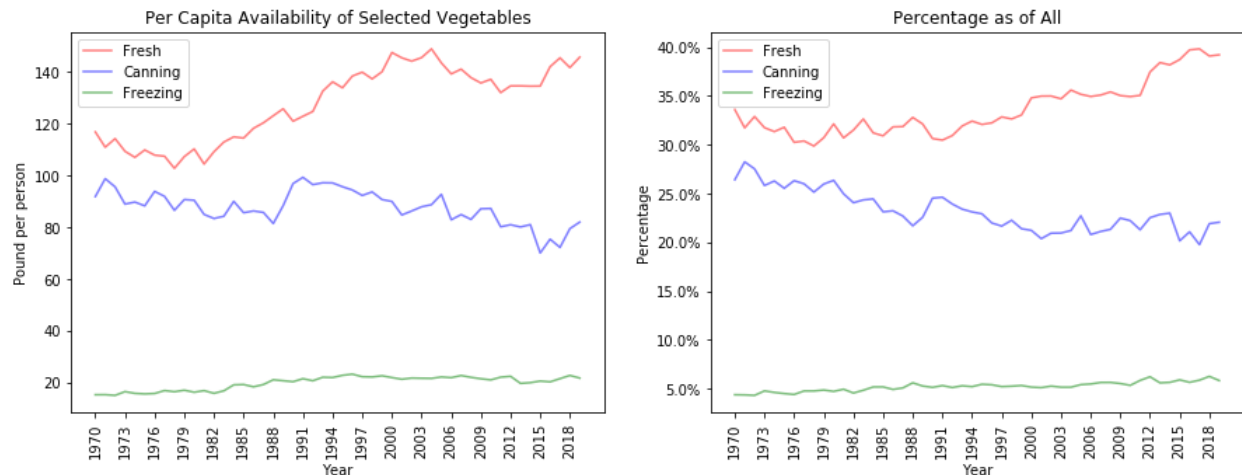


Figure: V3

Going into more details about the Fresh category, we analyzed the top 5 vegetable types for this category. Potatoes actually consisted of over 50% of the Fresh category in 1970 and the ratio has been decreasing. The other top Fresh vegetable types have increased over the years, especially the “Other” type. The Other type is a diversified mix of various vegetables, such as brussels sprouts, eggplant, garlic and etc. The rise of this vegetable type represents the growing options of fresh vegetables for an average American.

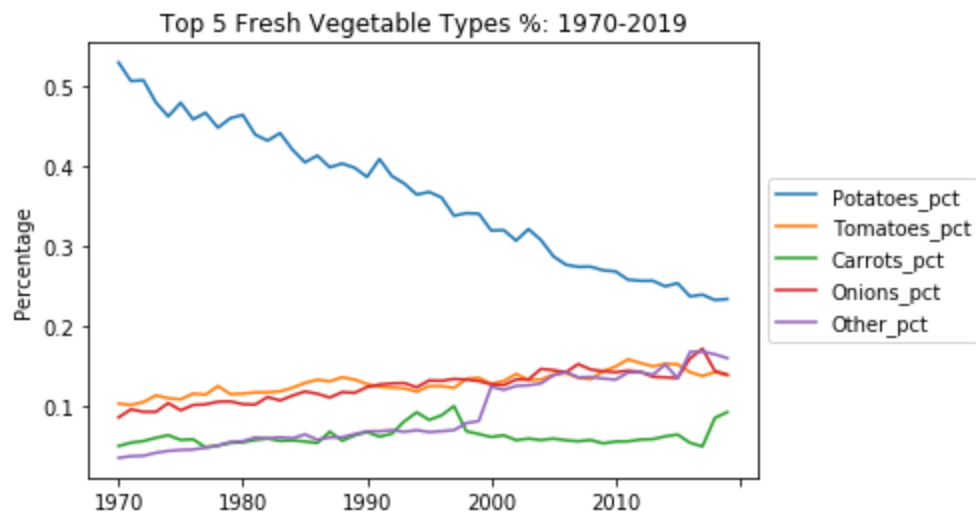


Figure: V4

Another interesting observation about the Canning category is that tomato is the dominant player, with an increasing weight of about 80% in 2019. The other vegetable types in the Canning category have either been decreasing or stayed flat.

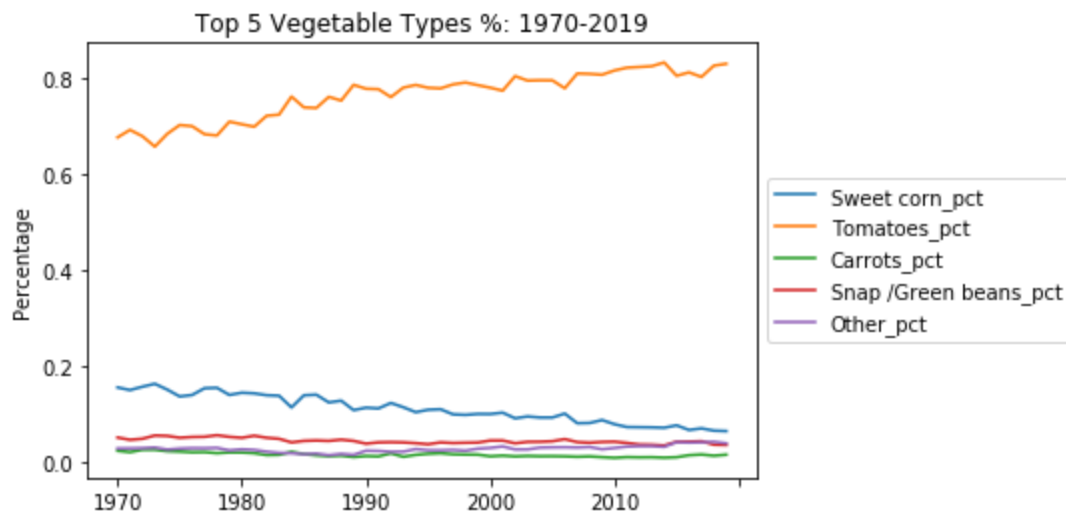


Figure: V5

Veggie: Supply and Utilization

We further broke down the domestic availability calculation for each fruit type and we selected a few interesting charts here. As we discussed in the previous sections, domestic availability is a function of domestic production, imports and exports. We showed two examples here for potatoes and tomatoes. As we can see in the charts below, the main source of fresh potatoes is domestic production, while the main source for tomatoes is actually imports. The domestic availability for tomatoes kept increasing over the years but production didn't keep up, and the increasing trend is actually supported by increasing imports.

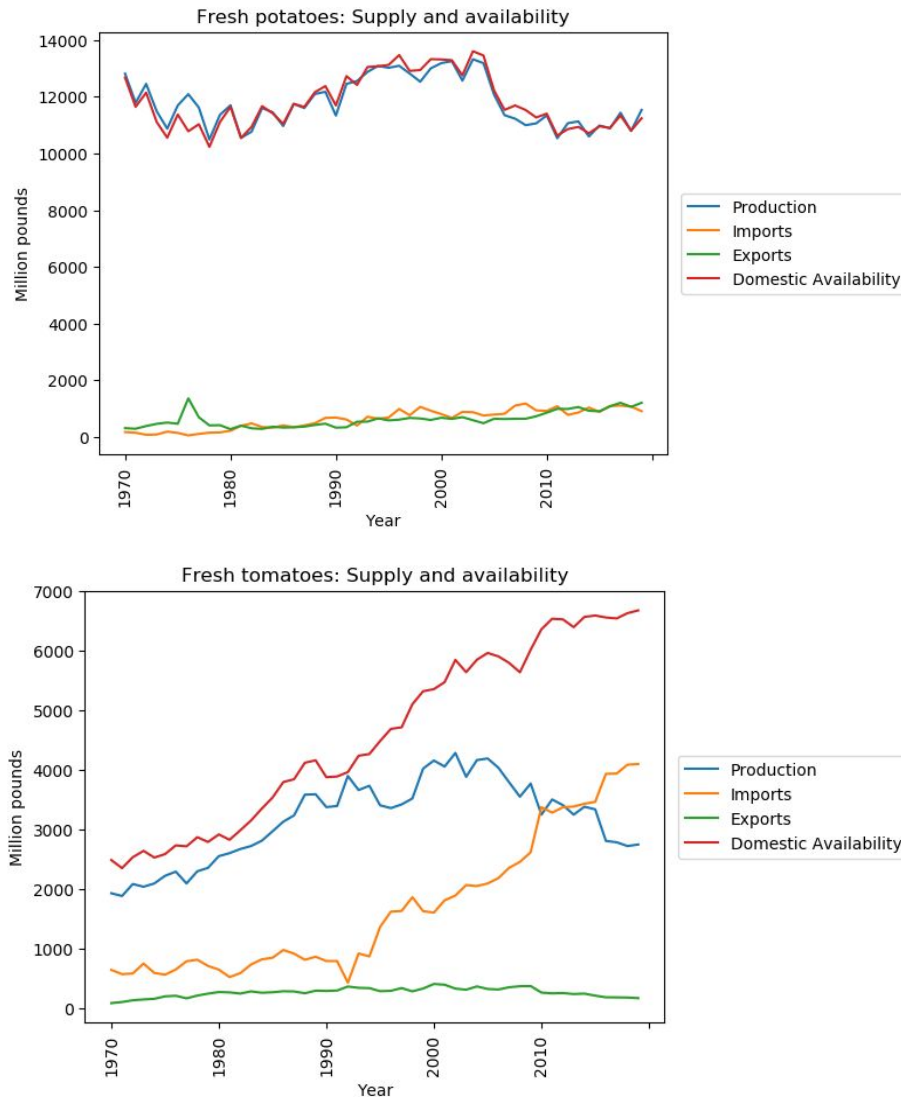


Figure: V6

Fruit

The analysis of fruit use, supply, etc is based on selected tables of the Fruit and Tree Nuts Yearbook from the United States Department of Agriculture(USDA). The original data is restored in the routine as *Project2_Singh_Zhang_Zhang\Fruit\Data_source All tables*, while the related code word is save in the ipynb files *Generak Fruit Table A.ipynb*, *Fruit Supply and Utilization G.ipynb* , respectively.

Fruit: Process Methods

First the total history use of selected fruits from 1976 to 2018 is explored here, as shown in the following diagram.

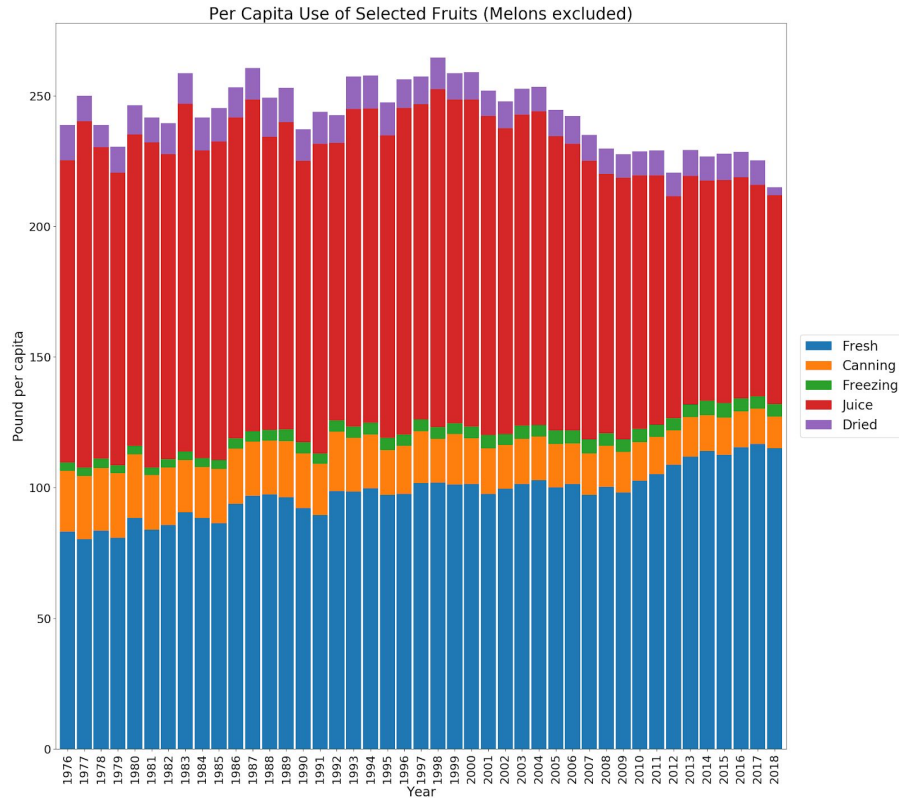


Figure X1

Overall, total per capita use of selected fruits in recent years (2012 - 2018) was a relatively low level, close to the value in the late 70s and early 80s. While, the peak of use of selected fruits was in the late 90s. About the fruit process methods, use of the fresh fruits kept increasing. Fresh fruit has replaced juice as the most use of fruit process type since 2010 (shown in plots as below). On the other hand, the juice use kept decreasing in recent years, however, it is the most required in old days.

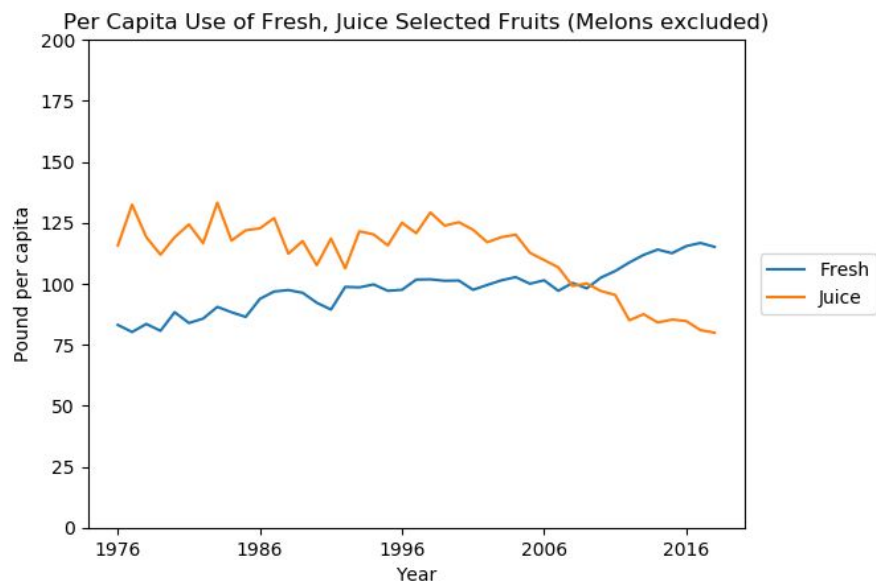


Figure F2

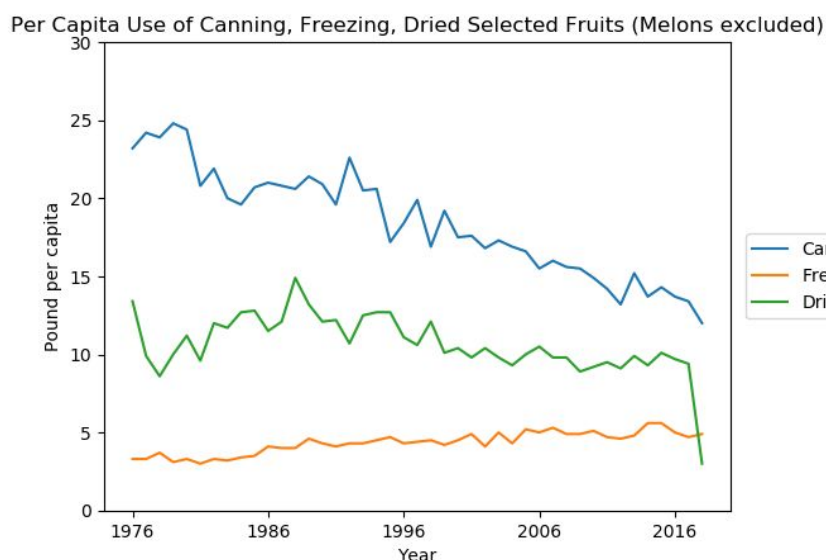


Figure F3

About the canning, freezing and dried type fruits, the use of these three types are much lower than the two types discussed above. Moreover, use of canning and dried kept decreasing. On the contrary, the use of freezing was gradually increasing.

Next, the use history will be explored according to the fruit types, aiming to determine the ranking of popularity.

Fruit: Types of Fruits

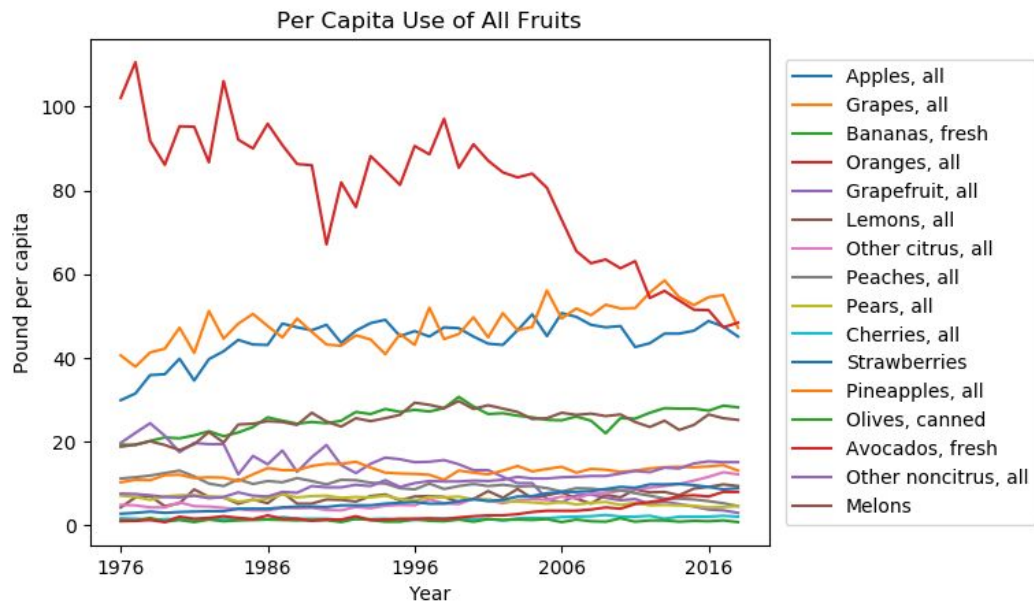


Figure F4

Due to vast types of fruits included in the plots, it is challenging to determine the popularity difference between each other. According to this diagram, oranges were the number one fruit type for the most time, however, its per capita use kept decreasing, especially for recent years. Moreover, apple and grape make up the second leading group in popularity.

To clarify more details about the fruit popularity ranking, a pie diagram was sketched according to the summary of per capita use from 1976 to 2018 for each fruit type (as below).

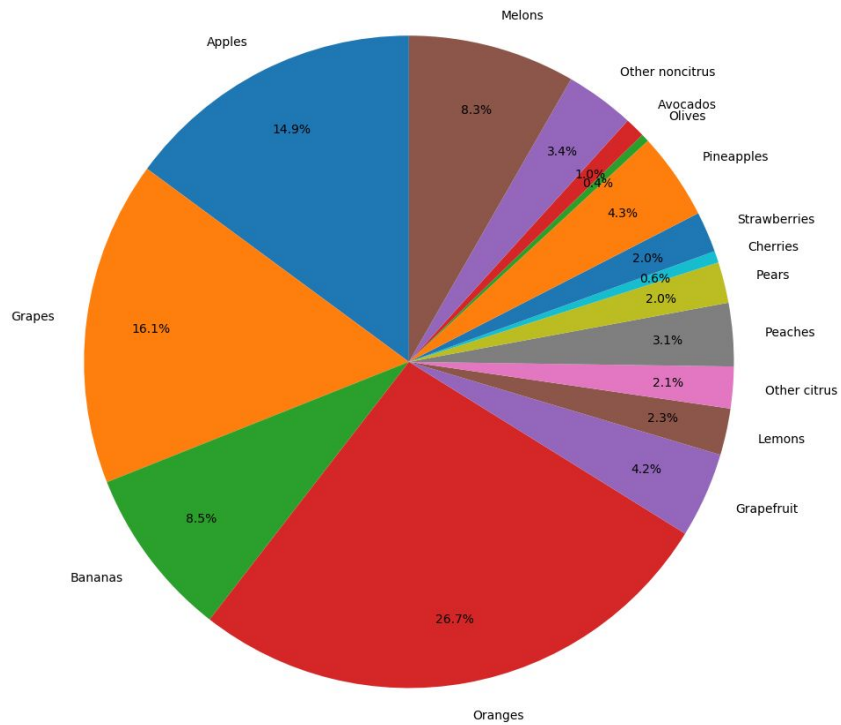


Figure F5

According to history per capita from the pie diagram above, the top five most popular fruits are Oranges(26.7%), Grapes(16.1%), Apples(14.9), Bananas(8.5%), Melons(8.3%) (followed by Pineapples, Grapefruits,etc.). The top five fruits contribute to about three quarters of the total history fruit use.

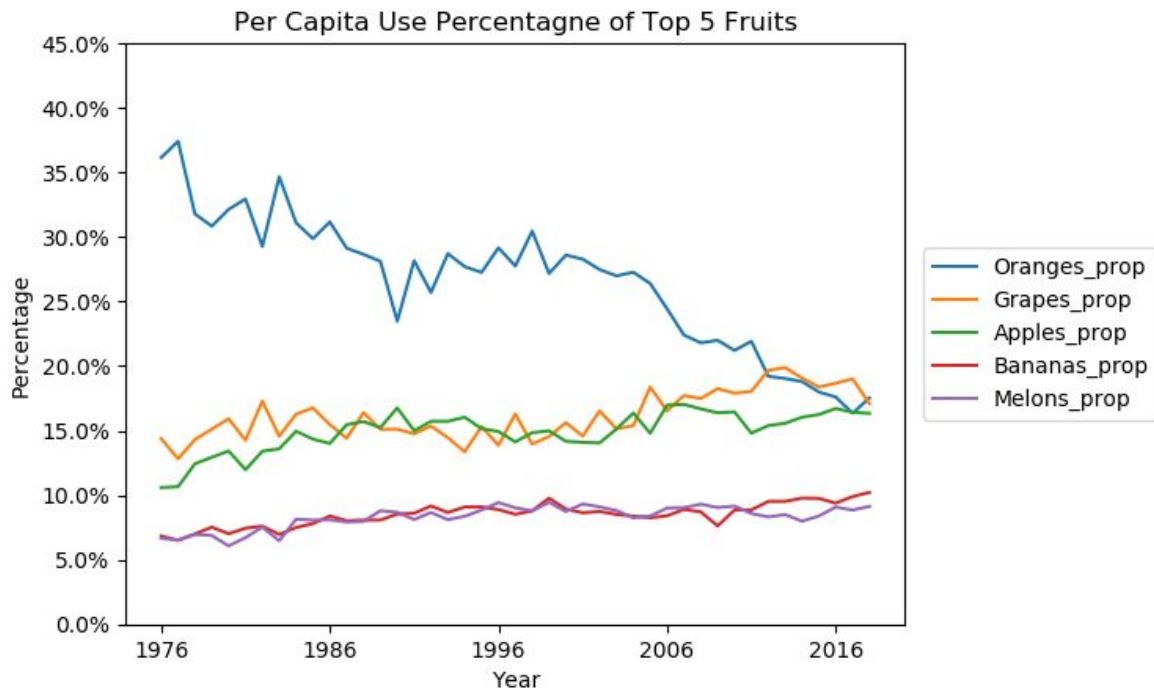


Figure F6

The trend of the proportion of top 5 fruits in the total fruits use is displayed here. Oranges were dominant before the 2000s, while grapes, apples were in the second leading group and bananas and melons were in the third leading group. In addition, grapes were the most popular fruits after 2012.

Fruit: Use of Top Fruits

In this section, the details about how the top popular fruits were used is explored. No details related to process methods other than fresh bananas are available. The assumption that the majority of bananas are fresh. Orange, grape, apple, melon and pineapple will be discussed as follows.

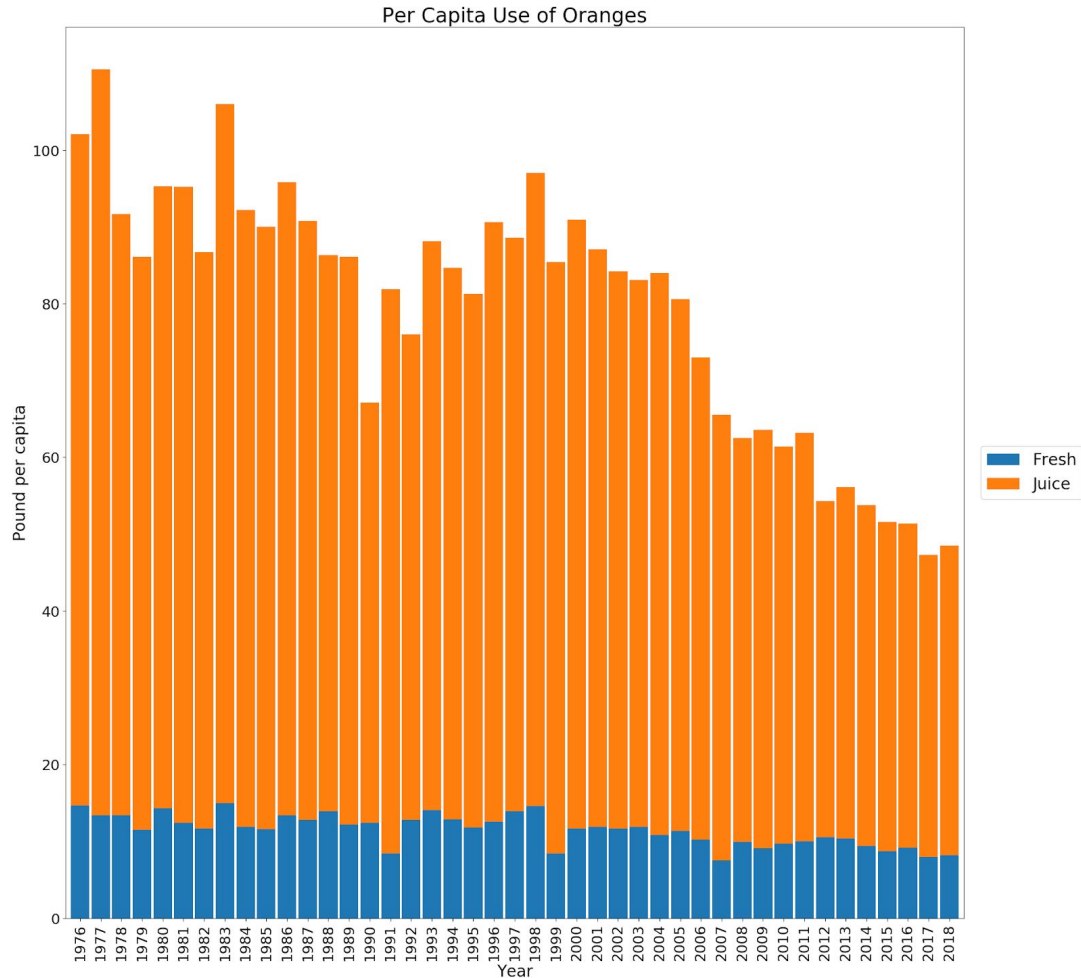


Figure F7

Americans used to consume a lot of oranges back to the 70s and 80s. The per capita use of orange peaked at 110.6 pounds in 1977. However, the orange use decreased rapidly after 2000.

The per capita use of orange was 48.4 pounds in 2018.

In addition, both fresh and juice use decreased for oranges. The per capita use of fresh orange was 8.2 pounds in 2018 with its peak value as 15 pounds in 1983. The per capita use of orange juice was 40.3 pounds in 2018 with its peak value as 97 pounds in 1977.

Regarding the proportion of fresh and juice use for oranges, orange juice contributed to over 80% of the total orange use over the years.

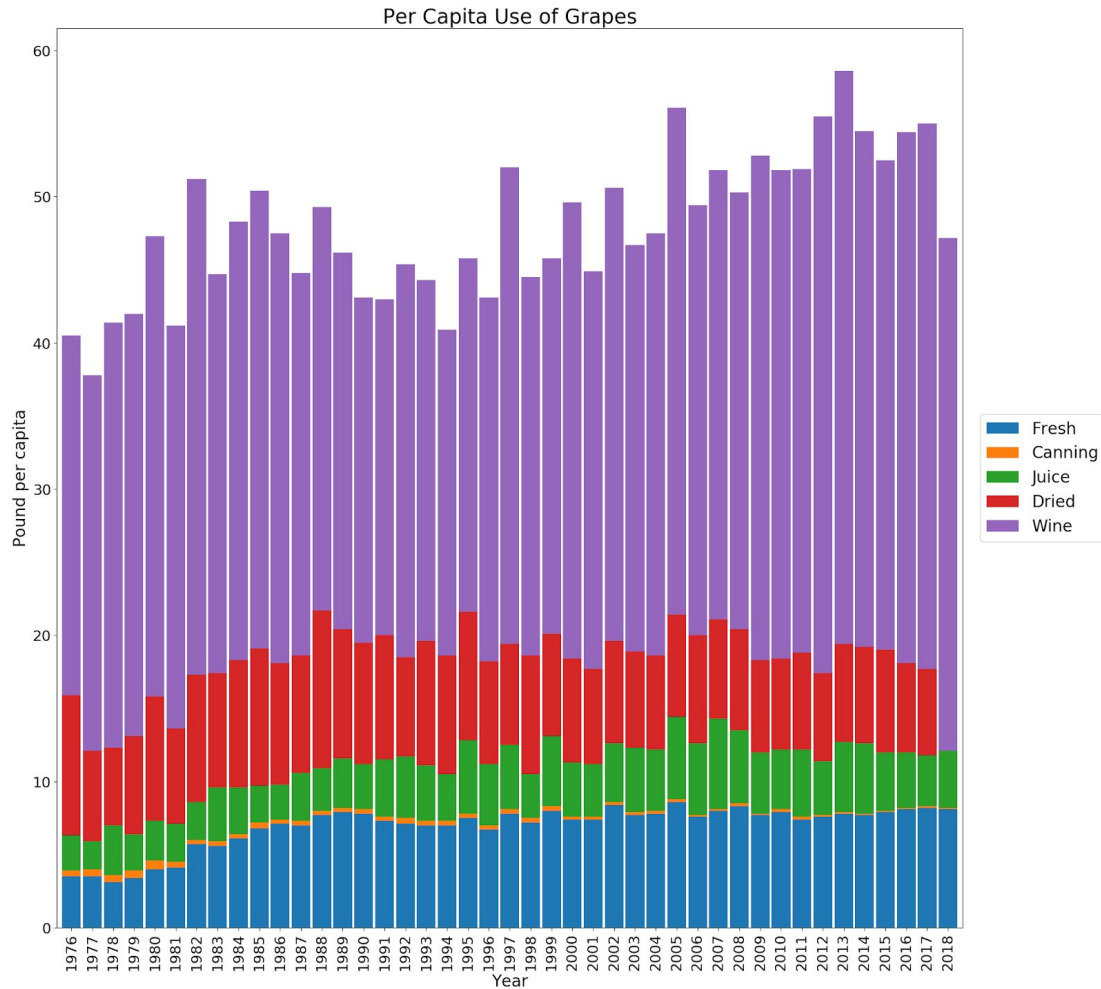


Figure F8

The per capita use of grapes was increasing steadily. Since the data of dried grapes discontinued, the use of grapes dropped in 2018. While, the total use was 55 pounds per capita in 2017, about 3.5 pounds lower than its peak value in 2013.

Regarding the proportion, wine contributed to over 60% of the total grape use most time. The share of fresh use increased from 8.6% in 1976 to 17.2% in 2018.

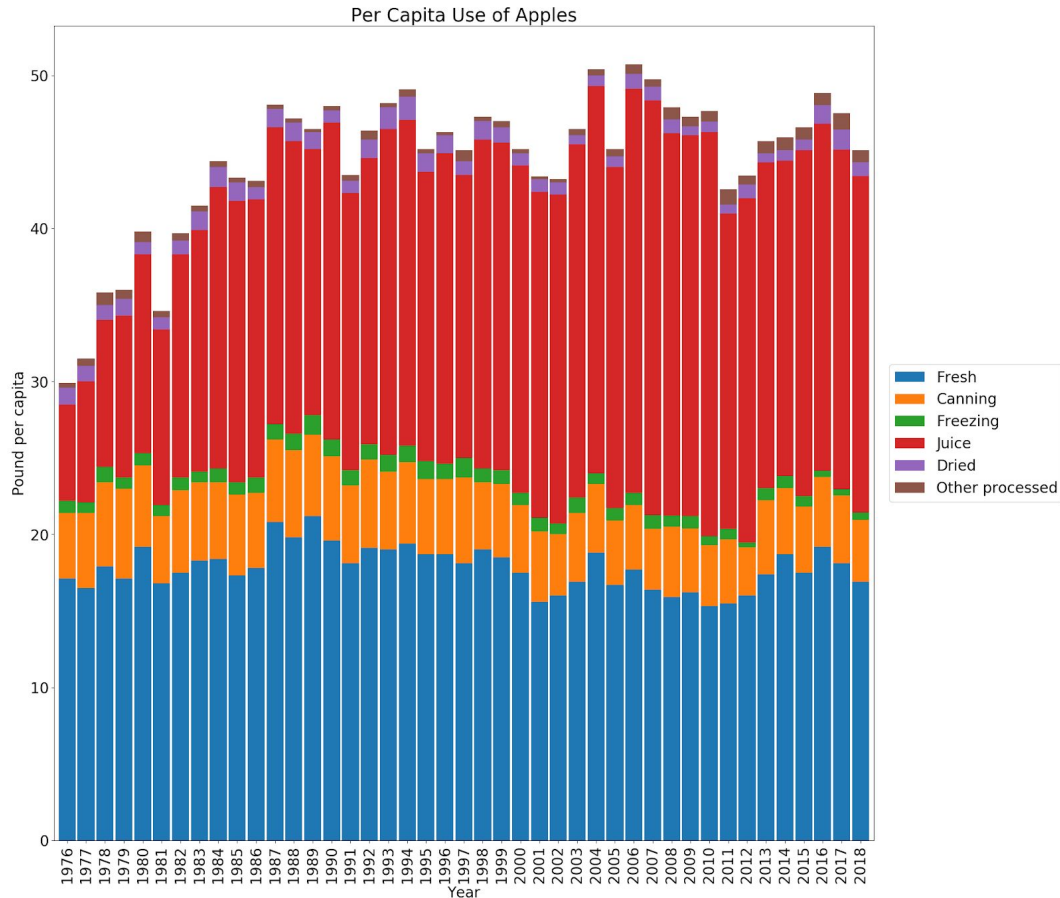


Figure F9

The per capita use of apples kept increasing these years. The total use was 29.9 pounds per capita at the lowest level in 1976, while its value was 45.1 pounds in 2018. In addition, the fresh apple use kept at a constant level across years (15.3 ~ 21.2 pounds), while the juice use increased from 6.6 pounds in 1976 to 22.0 pounds in 2018 with a peak value of 27.1 pounds in 2007.

Regarding the proportion, the share of juice use increased from 21%(1976) to 48%(2018). On the other hand, the share of fresh use decreased from 57%(1976) to 37%(2018).

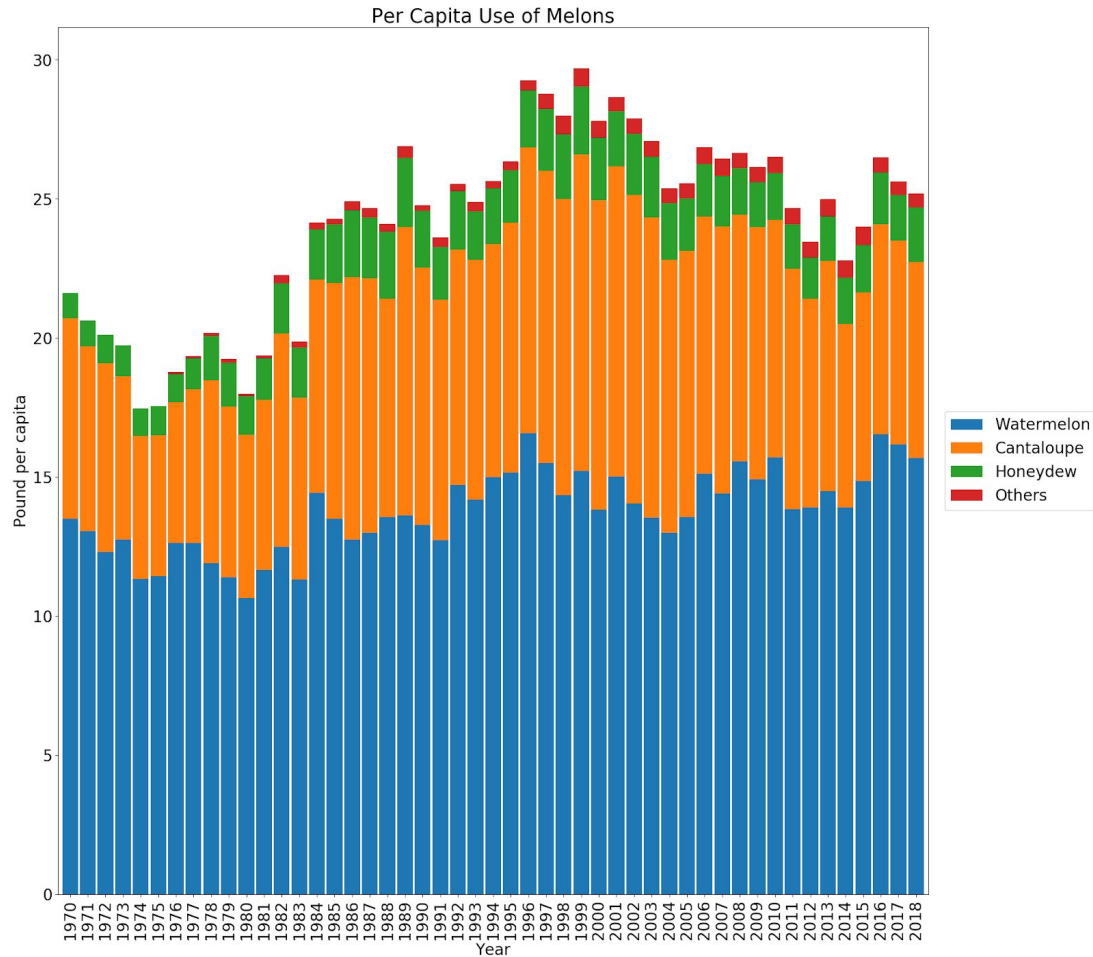


Figure F10

The per capita use of melons were pretty stable across years. Except for the time range from the early 70s to early 80s, the use of melos was above 20 pounds per capita in most years. The per capita use of melon was 25.2 pounds in 2018 with its peak value as 29.7 pounds in 1999.

Regarding the proportion, watermelons and cantaloupes made up over 90% of the total melon use.

Fruit: Supply and Utilization

In this section, the supply and utilization of top fresh fruits are explored. Production, imports, exports and domestic use of each fruit are plotted in the unit of million pounds versus year. Correlation maps about the four categories are also displayed as below.

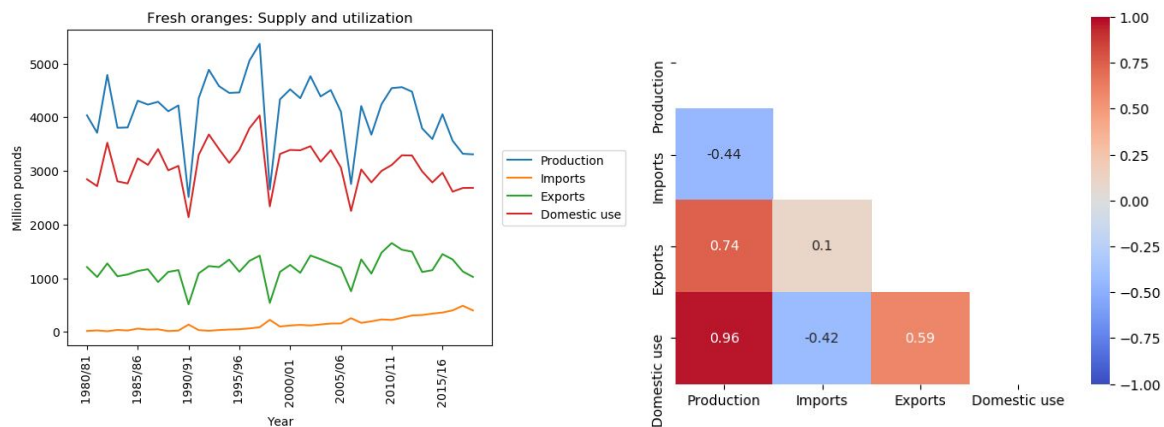


Figure F11

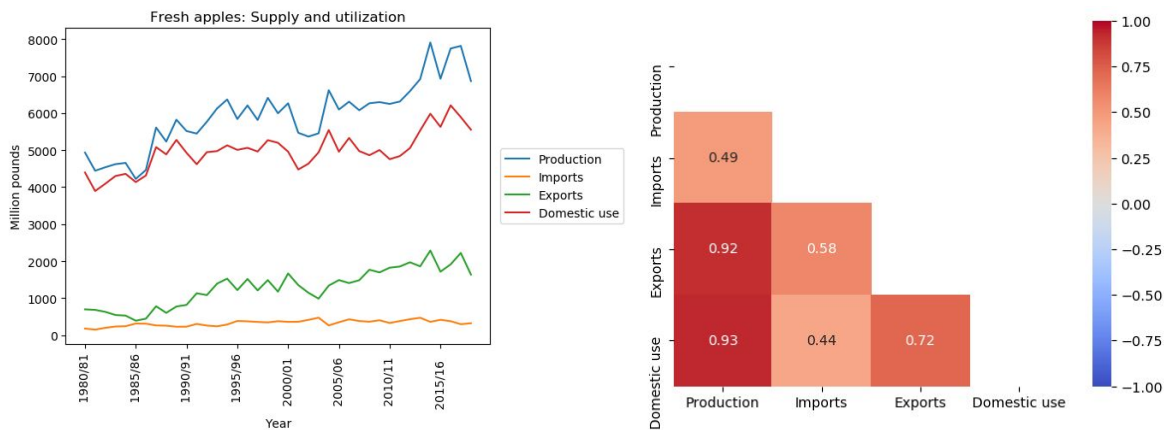


Figure F12

For oranges and apples, the US production is sufficient to satisfy the domestic use. At the same time, there is plenty of extra production to afford the exports. Imports proportion is relatively low for both.

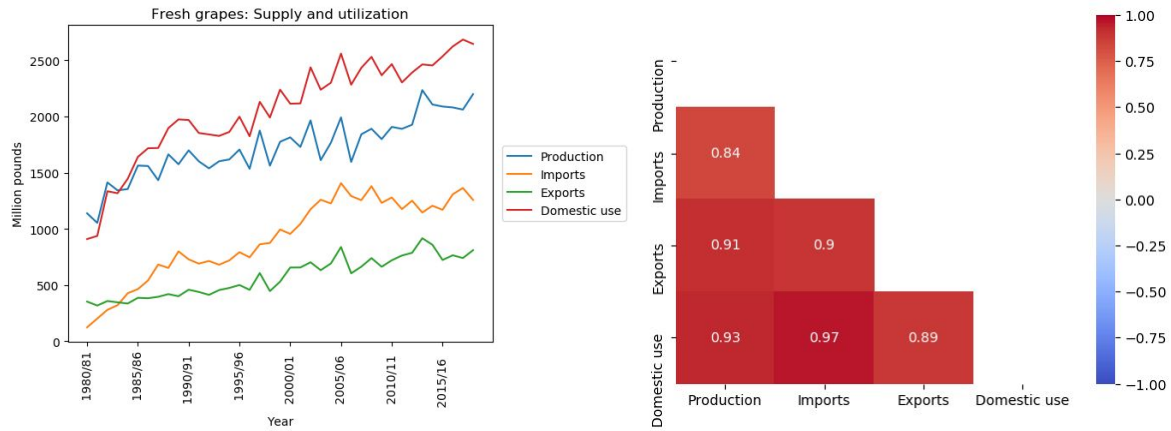


Figure F12

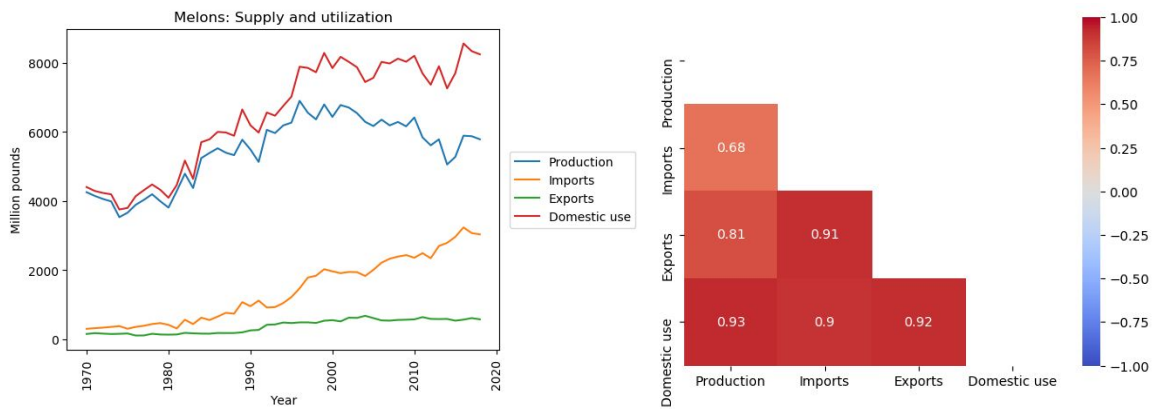


Figure F13

For grapes and melons, the US production is insufficient to satisfy the domestic use. With the increasing domestic use demand, the imports keep increasing. Especially for melons, the production has been decreasing in recent years.

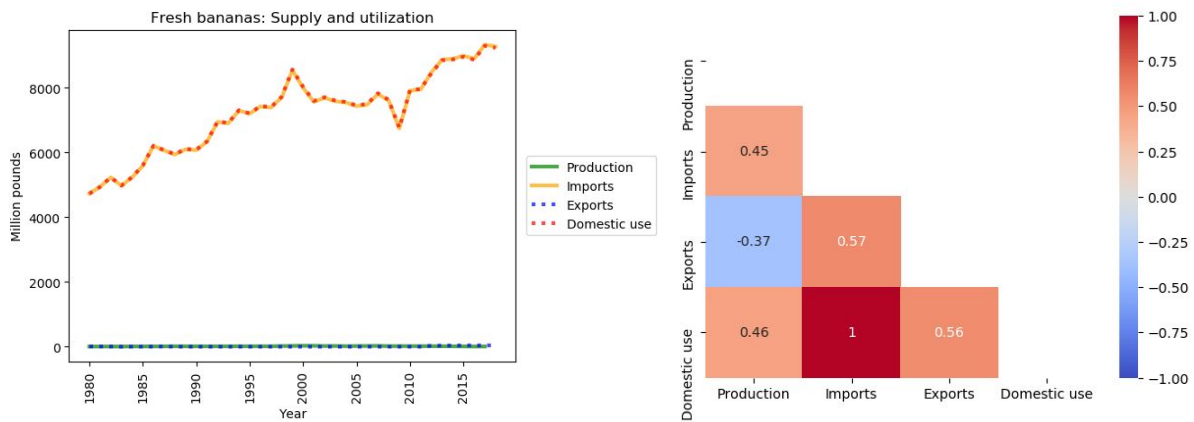


Figure F14

For bananas, over 99% of the domestic use depends on imports. The amount of local production is neglectable. According to [worldstopexports.com](http://www.worldstopexports.com) (<http://www.worldstopexports.com/bananas-imports-by-country/>), United States imported bananas totaled \$2,756,349,000 in 2019(17.6% of total imported bananas worldwide).

Meat: Supply and Utilization

We began by looking at the meat section initially from the consumption point of view. . First we looked at the meat consumption statistics across four important meat types consumed in the US - Beef, Pork, Lamb and Chicken. Figure M1 below shows the meat consumption statistics from 2005 to 2019. We can clearly see that the meat consumption for chicken has steadily increased in the past one decade. Some increase in consumption is also seen for Pork. Consumption for Beef has remained somewhat steady. Lamb represents a very small percentage of overall meat consumption in the US.

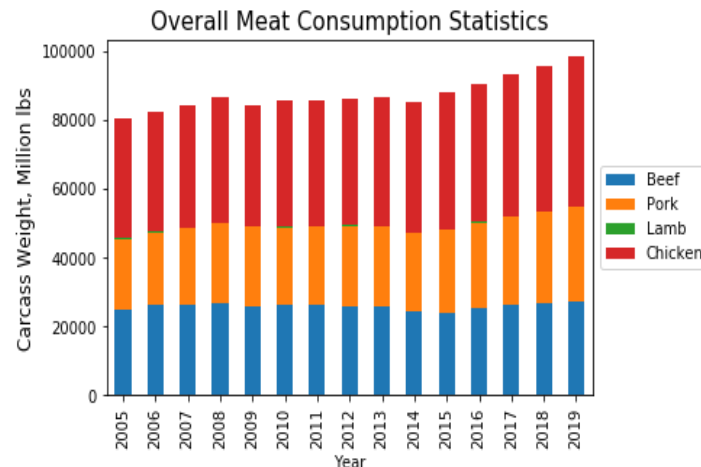


Figure: M1

We further looked at the import and export demands for each of these meat types to check to see how the meat consumption is affecting the import and export businesses. Figures M3 and M3 below show the import and export demands for these four meat types from 2005 to 2019. For figure M3, we can see that overall meat import initially declined from 2005 to 2011, then began to increase again from 2011 to 2015 and has been steady since the last five years. Figure M3 shows the export data for these four meat types from 2005 to 2019. It can be clearly seen that there has been a steady increase in overall meat export in these last 15 years. Chicken and Pork have led the export business with the demand for chicken rising much more than any other meat type. From figures M2 and M3, it can also be seen that the export and import for beef has relatively stabilized. As we would see in later sections, beef seems to become less popular in terms of per capita consumption and hence this data behavior is expected.

Figure M4 shows the top 10 export destinations for chicken between 2015 and 2016 - it can clearly be seen that while Mexico and Canada are top destinations for chicken, Cuba is leading the overall increase in demand. Figure M5 also shows the top 10 export destinations for Pork and we can clearly see that while Mexico and Japan lead the overall export, China has emerged as the top country with most increase in demand (by over 97%).

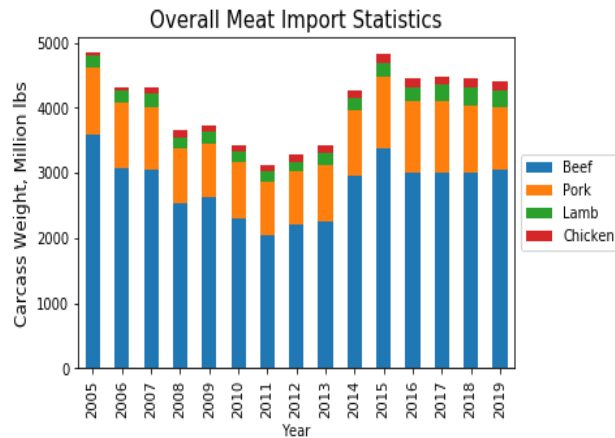


Figure M2

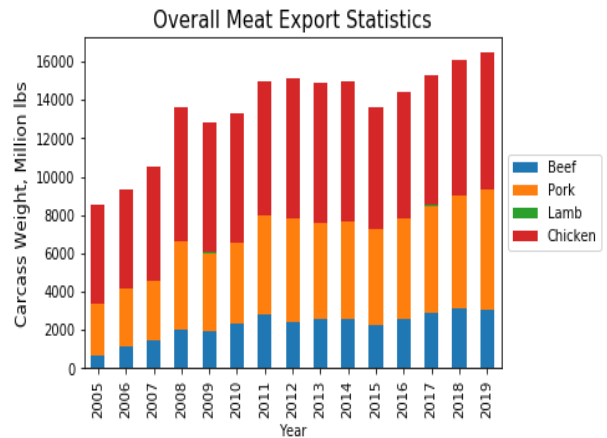


Figure M3

Top 10 U.S. Export Destinations For Chicken				
World Total	6,320,942,602	6,646,985,805	5.16%	
Rank	Country	2015	2016	Change
1	Mexico	1,472,970,894	1,405,191,036	-4.60%
2	Canada	399,633,625	357,898,849	-10.44%
3	Cuba	233,554,681	351,386,429	50.45%
4	Hong Kong	292,273,445	334,622,565	14.49%
5	Taiwan	384,796,157	333,114,456	-13.43%
6	Guatemala	229,656,267	262,771,055	14.42%
7	Angola	310,251,725	250,625,935	-19.22%
8	Philippines	158,093,303	200,944,335	27.10%
9	Iraq	161,646,472	185,588,197	14.81%
10	Kazakhstan	144,300,239	183,074,553	26.87%

Figure M4

Top 10 U.S. Export Destinations For Pork				
World Total	5,008,886,895	5,233,024,611	4.28%	
Rank	Country	2015	2016	Change
1	Mexico	1,553,429,901	1,603,909,506	3.25%
2	Japan	1,274,224,611	1,212,006,649	-4.88%
3	Canada	528,020,383	536,508,244	1.61%
4	China	251,539,974	497,504,751	97.78%
5	South Korea	493,908,573	406,726,148	-17.65%
6	Australia	176,406,737	186,968,634	5.99%
7	Hong Kong	125,918,339	148,185,713	17.68%
8	Colombia	109,530,466	119,350,687	8.97%
9	Philippines	72,043,356	75,603,810	4.94%
10	Dom. Republic	65,731,209	74,865,418	13.90%

Figure M5

Figure M6 below shows the wholesale price variations for particular categories of these 4 meat types (boxed beef choice, Pork Bellies and etc) over the last 2 decades. As we can see, lamb remains one of the priciest meat - with its price increasing steadily. This is also one of the reasons why lamb remains one of the lowest consumed meats in the US (Fig. M1). Price for Pork and Chicken seems to be much lower and on a decreasing trend - this enables higher consumption and import/exports of these meats (Fig. M1,M2 and M3)

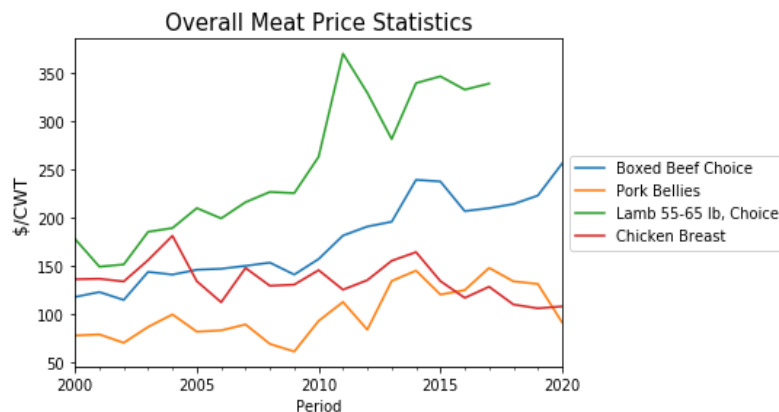


Figure M6

Meat: Pork and Chicken

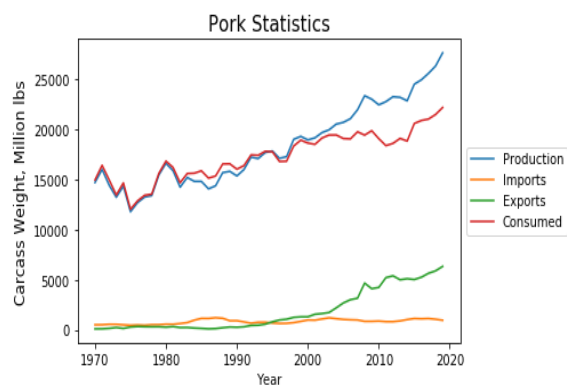


Figure M7

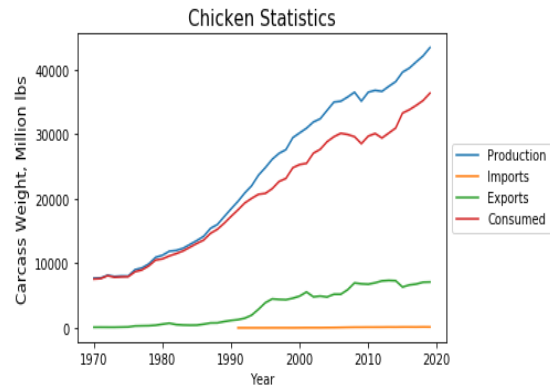


Figure M8

Pork and Chicken show very similar statistics in terms of Production, Consumption, Export and Import data. Chicken has shown a much higher production rate since 1990 and production exceeds consumption by ~10k million lbs. This enables higher export and larger price reduction for chicken compared to Pork overall (Fig. M6). In figure M9, we can see that the price for Chicken shows negative correlation (-0.45) with Production and Consumption. Thus, higher rates of production has enabled steep price reduction for chicken. We can also see that consumption shows high correlation (0.98) with production. Thus, higher production rate for chicken also enabled a higher rate of consumption in the US.

Pork has shown significant increase in production and export since 2005. This is largely attributed to higher pork export demands from countries like China and Mexico (Fig. M5). In figure M10 we can see that Pork export is highly correlated (0.94) with production. This shows the increasing production has enabled high export for Pork and thus the rate of its export exceeds the increase in rate of its consumption.

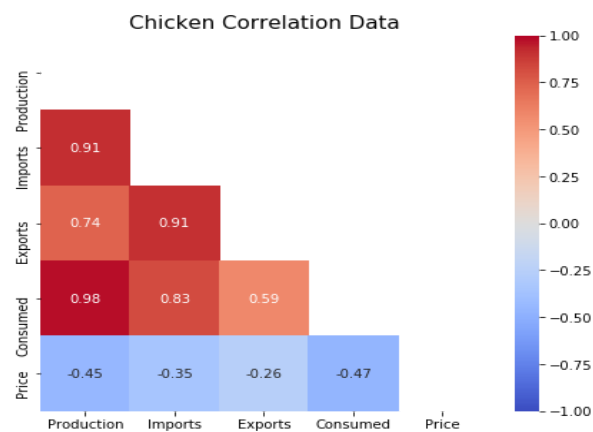


Figure M9

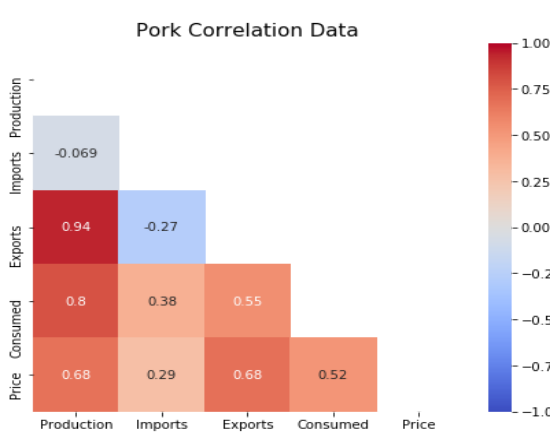


Figure M10

Meat: Lamb and Beef

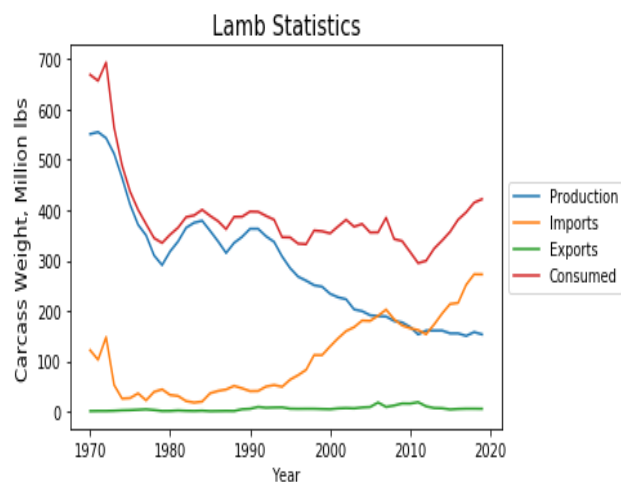


Figure M11

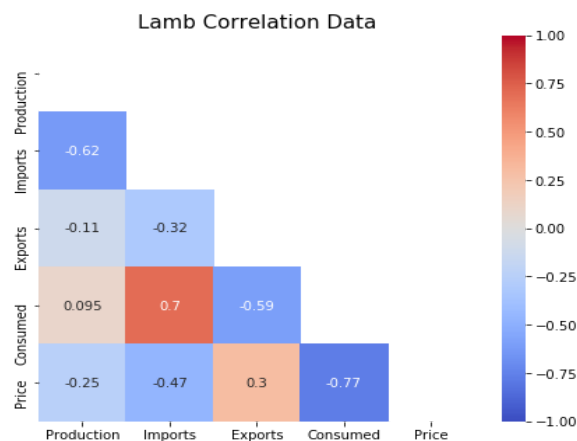


Figure M12

Figure M11 shows the statistics for lamb production, consumption, import and export from 1970 to 2020. It is interesting to note that both production and consumption for lamb decreased sharply from 1990 to 2000. Between 2000 and 2010 - the consumption of lamb remained kind of stable but started rising sharply from 2010 to 2020. Though there is this sharp increase in lamb consumption in this last decade, the production continues to decline - this has resulted in sharp increase in import from 2010 to 2020. As can also be seen from fig M12, consumption shows high correlation (0.7) with imports for lamb.

Figure M13 shows the beef statistics over the last 5 decades. It is interesting to observe the Production, Consumption, Import and Export for beef has remained relatively stable over the past decades. Also, since the population of the US has been increasing - without much increase in consumption for beef - the per capita consumption of beef has seen a decreasing trend since 1970 (fig. M14).

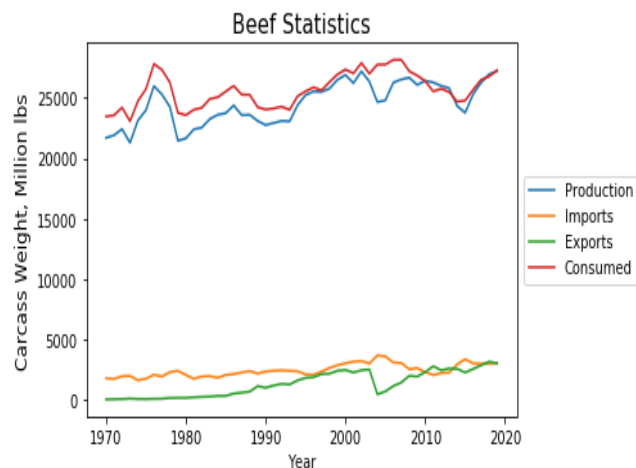


Figure M13

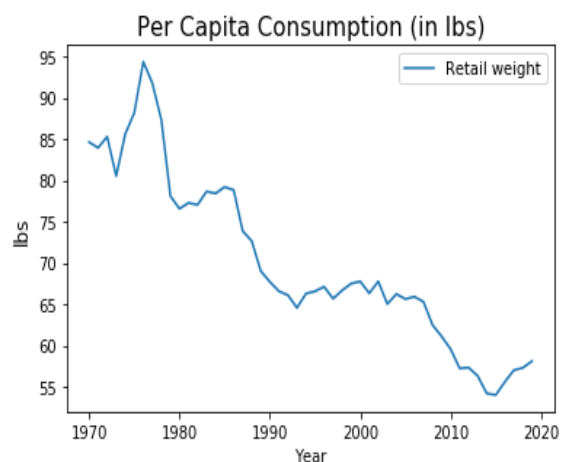


Figure M14

Dairy Consumption

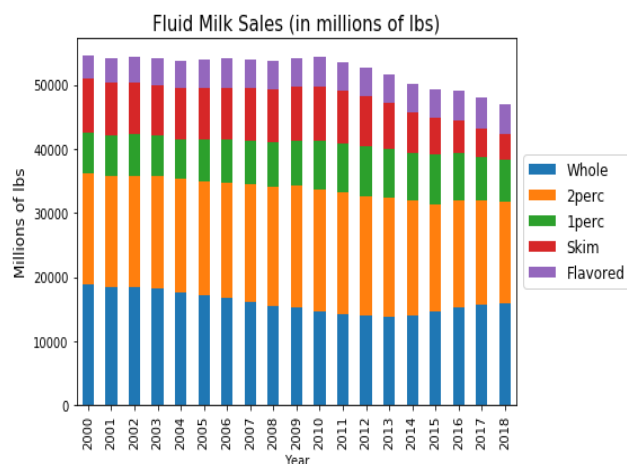


Figure M15

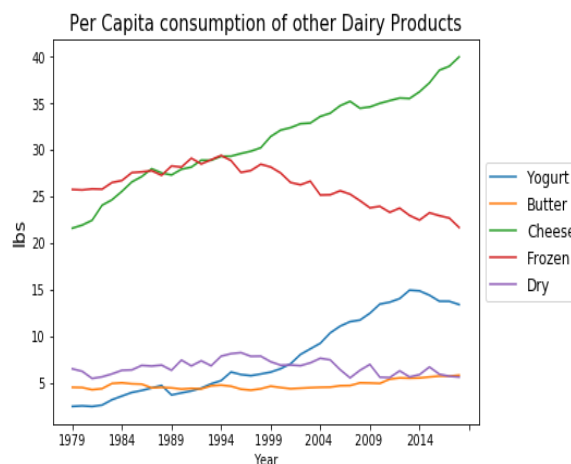


Figure M16

Fig. M15 shows the fluid milk sales in the US over the past 2 decades. Few interesting trends to note are:

- a) Overall milk sales remained stable from 2000 to 2010 but have been on a decreasing trend since 2010. This shows that since the past two decades, per-capita milk consumption has steadily decreased in US
- b) Relative consumption of whole milk initially decreased from 2000 to 2012 but has been on an increasing trend since 2012. Consumption for other categories like 2% fat, 1% fat, flavored and skim have decreased a little since 2012.

Figure M16 shows the per-capita consumption of other dairy products since 1979. Cheese remains the highest consumed product since 1990 and continues to see a sharp increase in last decade. Figure M17 shows the relative consumption of different cheese types - Cheddar and Mozzarella being the highest consumed ones. Though Frozen products remain the second most consumed product, they have seen a steady decline in consumption since 1990. Demand for Yogurt seems to be increasing steadily since 2004 but has become relatively stable since 2014.

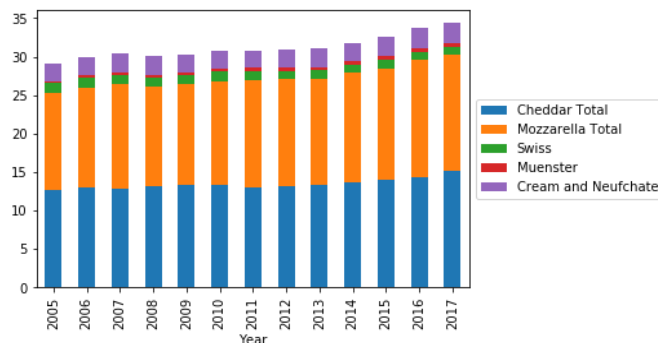


Figure M17

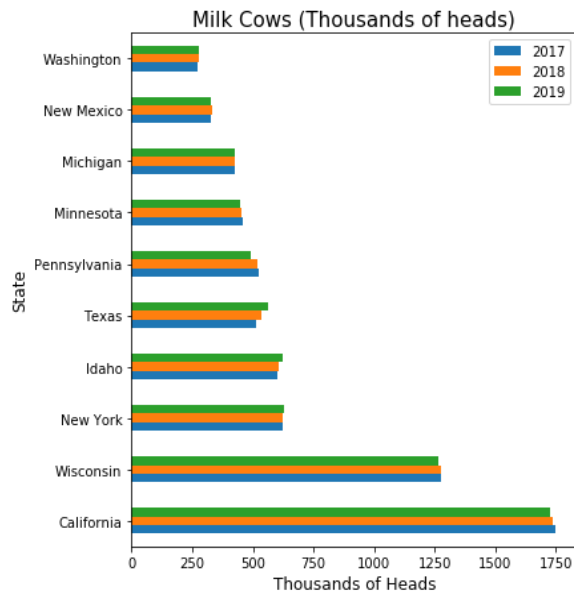


Figure M18

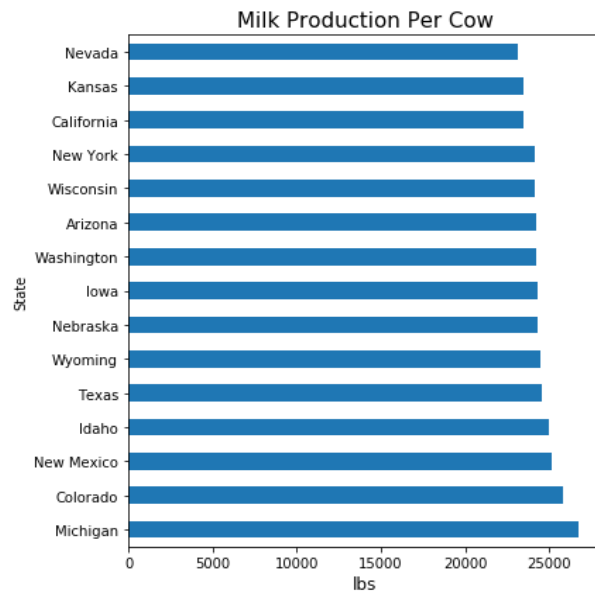


Figure M19

Figure M18 shows the top 10 US states with the highest number of milk cows between 2017 to 2019. As we can see California and Wisconsin have the highest number of milk cows. Figure M19 shows the top 10 US states with highest milk production per cow. Michigan and Colorado lead the race with highest milk production per cow.

Figure M20 shows top 10 US states with the highest total milk production in the US in 2019. As we can clearly notice from fig. M18 and M19 - since California and Wisconsin have the highest number of milk cows - they also lead the total milk production in Fig. M20. As we can also see from Fig. M19, Michigan and Colorado had the highest milk production per cow - but a much lesser number of milk cows (Fig. M18) and therefore lay behind in the race of total milk production (Fig. M20).

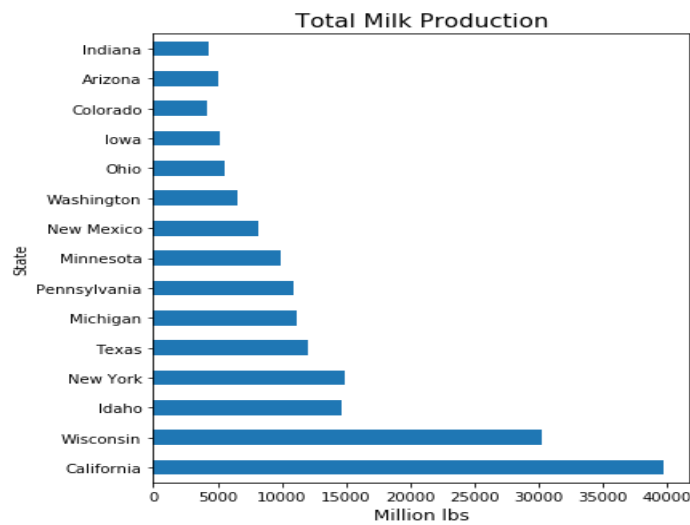


Figure M20