research

Gui Gao, Ishita Khanna, Shu Wang, Chen Wanxin 2022-05-20

1 Introduction

With the improvement of people's living standards, more and more people begin to pay attention to diet. As we all know, a healthy diet can help us maintain a healthy body and figure, and effectively prevent dangerous diseases such as high blood pressure and high blood lipids. In our daily diet, our calorie intake mainly comes from three major nutrients, namely protein, fat and carbohydrates. Studies have shown that the proportion of the three major nutrients in the daily diet plays a very important role in human health. On the other hand, people in different regions have different eating habits due to differences in climate and terrain.

In this report, we research the proportion of various food groups in the Australian diet changed over time, and research different food items FAO concentration in Australia and at different period of time. So that we can see whether people are shifting towards being health conscious or are they taking a healthy and balanced diet or not? Besides, we research the differences in diet between the United States and Japan and analyze the changes in the per capita calorie intake and the intake of the three major nutrients in the United States and Japan since 1960, starting from two aspects of eating habits and time trends A comparative analysis of the eating habits and health of the two countries. Rice consumption vs. latitude and region question is also related to our diet.

2 Analysis1

```
# read data
daily_caloric_supply <- read_csv("Data/daily-caloric-supply-derived-from-carbohydr
ates-protein-and-fat.csv")
dietary_compositions <- read_csv("Data/dietary-compositions-by-commodity-group.csv")
overweight_calories <- read_csv("Data/share-of-adult-men-overweight-or-obese-vs-da
ily-supply-of-calories.csv")</pre>
```

```
# data wrangling
daily_caloric_supply <- daily_caloric_supply %>%
  rename_all(str_remove, pattern = "\\(FAO.+\\)") %>%
  select(-Code)

dietary_compositions <- dietary_compositions %>%
  rename_all(str_remove, pattern = "\\(FAO.+\\)") %>%
  select(-Code)
```

2.1 Research Question1

How the proportion of various food groups in the Australian diet changed over time?

```
# data figure1
figure1 <- dietary_compositions %>%
  filter(Entity == 'Australia') %>%
  pivot longer(cols = -c(Entity, Year),
               names_to = 'Variable',
               values_to = 'Value') %>%
  ggplot(aes(x = Year,
             y = Value,
             fill = Variable)) +
  geom_area(color = 'white') +
  scale fill viridis d() +
  labs(y = 'Kilocalories per Person per Day',
       title = 'Kilocalories per Person per Day in Australia') +
  theme classic() +
  theme(plot.title = element_text(hjust = 0.5),
        legend.position = 'bottom',
        text = element text(size = 8)) +
  transition_reveal(Year)
figure1
animate(figure1,
        res = 300,
        width = 2000,
        height = 1125,
        renderer = gifski_renderer())
```

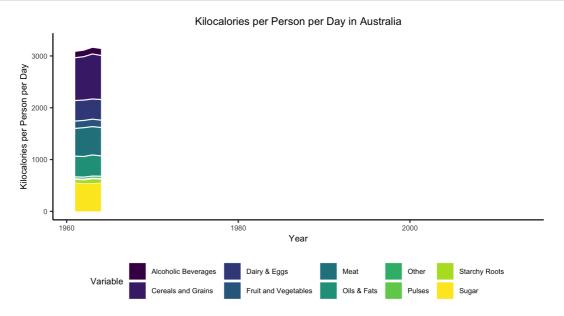


Figure 2.1: Kilocalories per Person per Day in Australia

```
anim_save("figure1.gif")
```

2.2 Data Explanation

According to Figure 2.1, we can see the different colours represent different food groups and the larger the area, the greater the proportion of the Australian diet. It is clear that they prefer cereals and grains, meat, fats and sugary foods to pulses and starchy roots.

2.3 Research Question2

How the calories from animal protein varies around the world?

```
# read world map data
world <- readOGR(dsn = 'World_Countries_(Generalized)/.')</pre>
```

```
OGR data source with driver: ESRI Shapefile
Source: "/Users/gulgul/Desktop/assignment4-groupl/World_Countries_(Generalized)",
layer: "World_Countries__Generalized_"
with 249 features
It has 7 fields
```

```
world shp <- world %>%
  st as sf()
world_data <- daily_caloric_supply %>%
  select(Entity, Year, `Calories from animal protein `) %>%
  merge(world_shp,
        by.x = "Entity",
        by \cdot y = "COUNTRY") %>%
  st_as_sf()
# data figure2
figure2 <- world data %>%
  ggplot(aes(fill = `Calories from animal protein `)) +
  geom_sf(colour = NA) +
  labs(x = 'Longitude',
       y = 'Latitude',
       title = ' Year: {closest_state}') +
  scale_fill_viridis_c(na.value = 'grey') +
  theme_void() +
  theme(legend.position = 'bottom') +
  transition states(states = Year)
figure2
animate(figure2,
        res = 300,
        width = 2000,
        height = 1125)
```

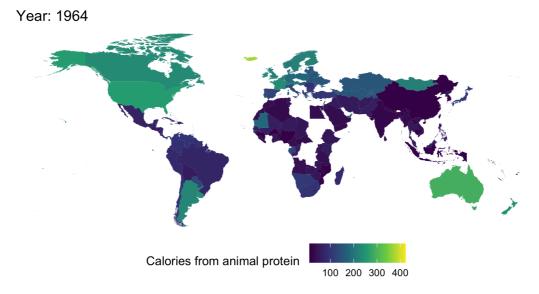


Figure 2.2: Year: {closest_state}

anim_save('figure2.gif')

2.4 Data Explanation

To better represent the variation in calories provided by animal protein around the world, I first downloaded data from a world map online, then matched it to the dataset I chose by country name, then filled in the colours according to the calories provided by animal protein to create Figure 2.2. The closer the color is to green, the more calories from animal protein, and conversely the closer the color is to blue the less there is. Overall, the amount of calories from animal protein has increased worldwide.

2.5 Reference

Observing the Figure 2.2 we can find that people living in North America, Oceania, and Europe consume more animal protein to provide calories, simply put, their diet composition prefers meat products, but also from the side to reflect the continued high consumption of livestock products in almost all developed countries Stoll-Kleemann and O'Riordan (2015).

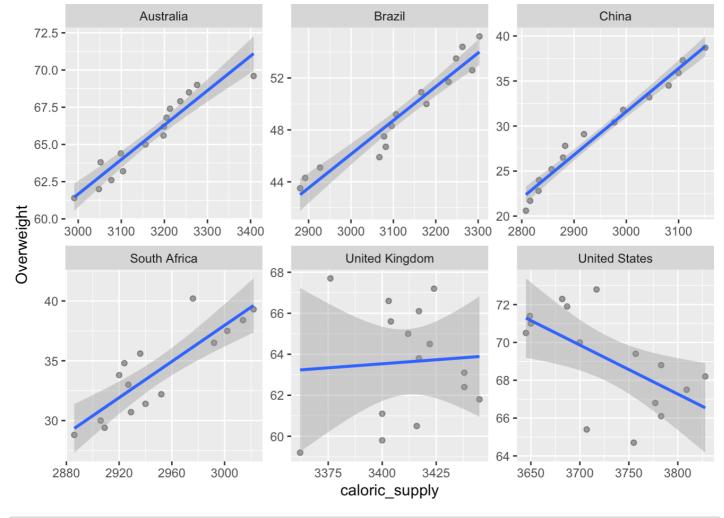
2.6 Research Question3

In Australia, Brazil, China, South Africa, United Kingdom and United States, which country has the relatively best linear model of the relationship between overweight or obese and caloric supply since 2000?

```
# linear models
by_entity2 <- by_entity %>%
  group_by(Entity)%>%
  nest()

fit_lm <- function(x){
  lm(Overweight~caloric_supply, data = x)
}
mapped_lm <- map(by_entity$data, fit_lm)</pre>
```

```
scatter_plot <-
ggplot(data = by_entity,
    aes(x = caloric_supply,
        y = Overweight)) +
    geom_point(alpha = 0.4) +
    facet_wrap(~Entity, scales = "free") +
    geom_smooth(method = "lm")
scatter_plot</pre>
```



```
# A tibble: 6 \times 4
# Groups:
            Entity [6]
  Entity
                 data
                                    model
                                            tidy
  <chr>
                 st>
                                    t> <list>
1 Australia
                 <tibble [15 × 3]> <lm>
                                            <tibble [2 × 5]>
                                            <tibble [2 × 5]>
2 Brazil
                 <tibble [15 × 3]> <lm>
3 China
                 <tibble [15 × 3]> <lm>
                                            <tibble [2 × 5]>
4 South Africa
                 <tibble [15 × 3]> <lm>
                                            <tibble [2 \times 5]>
5 United Kingdom <tibble [15 \times 3]> <lm>
                                            <tibble [2 × 5]>
6 United States <tibble [15 × 3]> <lm>
                                            <tibble [2 × 5]>
```

```
entity_glance <- entity_model %>%
  mutate(glance = map(model, glance)) %>%
  unnest(glance) %>%
  select(Entity, r.squared, AIC, BIC)
```

Table 2.1: Goodness_of_fit_measures

Entity	r.squared	AIC	BIC
Australia	0.9079612	40.73470	42.85885
Brazil	0.9232400	48.68766	50.81181
China	0.9727043	46.42461	48.54876
South Africa	0.7375348	66.86584	68.98999
United Kingdom	0.0040535	77.89588	80.02003
United States	0.3628869	69.38632	71.51047

2.7 Data Explanation

The scatter plot shows that the Chinese fitted linear model is the best, while, according to Table2.1, we can find that Chinese linear model has a maximum r.squared value around 0.97 and there are relatively small AIC and BIC values of around 46.42 and 48.55 respectively. Therefore, we can find that Chinese people's overweight or obese are more affected by calorie intake.

2.8 Reference

R-squared is the percentage of outcome variable variation explained by the model, and describes how close the data are to the fitted regression. In general, the higher the R-squared value, the better the model fits. AIC and BIC both aim at achieving a compromise between model goodness of fit and model complexity. The preferred models are those with minimum AIC/BIC (Yang and Berdine (2015)).

3 Analysis2

```
# reading csv file
dietary_csv <- read.csv("Data/dietary-composition-by-country.csv")</pre>
```

3.1 Research Question1

How much FAO i.e. Fats Animal Oil is in Vegetable Oil in Australia that is consumed by people in different year?

```
# filter the data
country_vege_oils <- dietary_csv %>%
    filter(Entity == "Australia")
# selecting particular columns
selection <- country_vege_oils %>% select(Year, Vegetable.Oils..FAO..2017..)
# arranging in descending order based on Vegetable oil FAO
arrange(selection ,desc(Vegetable.Oils..FAO..2017..))
```

```
Year Vegetable.Oils..FAO..2017..
  2012
1
                                    569
2
  2013
                                    550
  2010
                                    547
3
4
  2011
                                    530
5
  2004
                                    524
6
  2009
                                    522
7
   2005
                                    516
   2006
                                    508
  2007
                                    488
10 2001
                                    479
11 2008
                                    479
12 1999
                                    459
13 2002
                                    450
14 2000
                                    441
15 1992
                                    428
16 1997
                                    427
17 1993
                                    426
18 2003
                                    426
19 1998
                                    418
20 1991
                                    403
```

21 1	1996	400
22	1994	398
23 1	1995	398
24 1	1990	365
25 1	1989	354
26	1987	335
27	1988	334
28	1986	311
29	1985	299
30	1980	288
31 1	1982	285
32	1983	285
33 1	1981	273
34	1979	265
35 1	1984	258
36	1977	232
37	1978	232
38	1975	188
39	1976	186
40	1974	181
41 1	1973	175
42	1972	167
43	1970	150
44	1971	136
45 1	1969	114
46	1966	113
47 1	1967	106
48 1	1968	105
49	1965	103
50 1	1964	100
51 1	1963	92
52	1961	78
53	1962	78

3.2 Research Question2

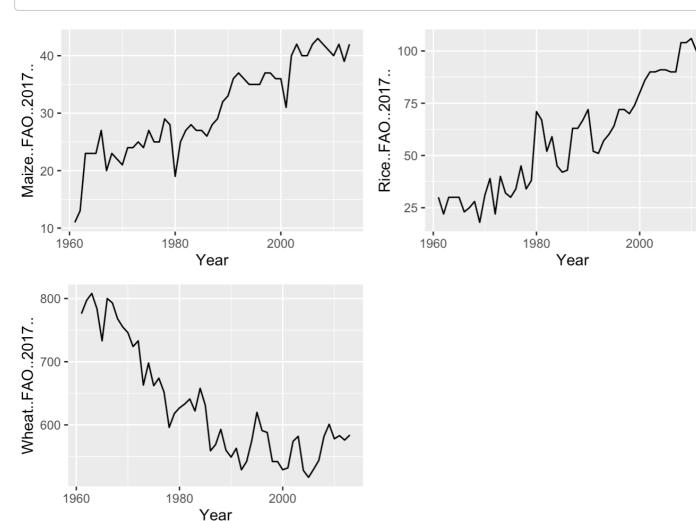
Comparing the FAO in maize, rice and wheat over the years in single figure to see that they all decreased, increased or differs?

```
# plotting Maize FAO on different years
maize_plot <- ggplot(country_vege_oils, aes(x = Year, y = Maize..FAO..2017..)) +
   geom_line()</pre>
```

```
# plotting Rice FAO on different years
rice_plot <- ggplot(country_vege_oils, aes(x = Year, y = Rice..FAO..2017..)) +
  geom_line()</pre>
```

```
# plotting Wheat FAO on different years
wheat_plot <- ggplot(country_vege_oils, aes(x = Year, y = Wheat..FAO..2017..)) +
   geom_line()</pre>
```

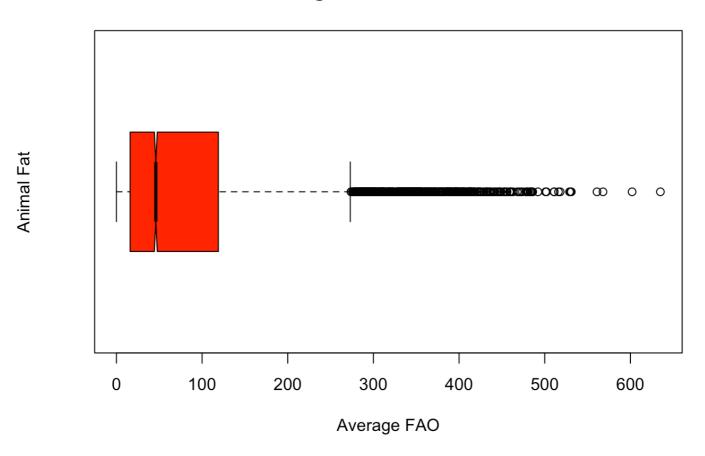
```
# joining three plots as one figure
ggarrange(maize_plot, rice_plot, wheat_plot)
```



3.3 Research Question3

Distribution of FAO in Animal Fat and Vegetable Oil against the averages and skewness.

Average FAO in Animal Fat



```
b$stats
```

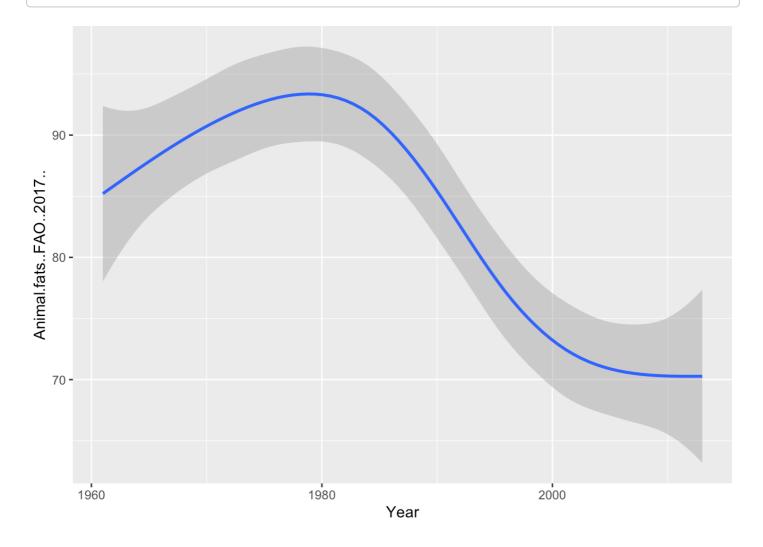
```
[,1]
[1,] 0
[2,] 16
[3,] 46
[4,] 119
[5,] 273
```

3.4 Research Question4

Finding the relation between two variables i.e. Year and FAO in Animal Fat.

```
Call:
lm(formula = Year ~ Animal.fats..FAO..2017.., data = dietary_csv)
Residuals:
     Min
               1Q
                   Median
                                3Q
                                        Max
-27.2921 -13.1134
                   0.1706 13.1531
                                   30.0798
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)
                         1.988e+03 2.144e-01 9273.059 < 2e-16 ***
                                               -7.498 7.08e-14 ***
Animal.fats..FAO..2017.. -1.276e-02 1.702e-03
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 15.28 on 8979 degrees of freedom
Multiple R-squared: 0.006223, Adjusted R-squared: 0.006112
F-statistic: 56.22 on 1 and 8979 DF, p-value: 7.078e-14
```

```
ggplot(dietary_lm) +
geom_smooth(aes(x=Year, y=Animal.fats..FAO..2017..))
```



3.5 Data Explanation

Maize and Rice FAO is higher in later years but the wheat growth becomes less in later years in Australia. Same as maize and rice, Vegetable FAO is growing in later years in Australia. There is no particular relation between year and Animal Fat because of different countries but it says that animal fat increases with the increase in year but gets low as well in some countries. So its fluctuating.

4 Analysis3

```
data <- read.csv("Data/daily-caloric-supply-derived-from-carbohydrates-protein-and
-fat.csv")

mydata <- data %>% filter(Entity %in% c("United States", "Japan"))
```

```
pct_miss(mydata) #0 missingness in the UK and Iceland data
pct_miss_case(mydata)
pct_miss_var(mydata)
```

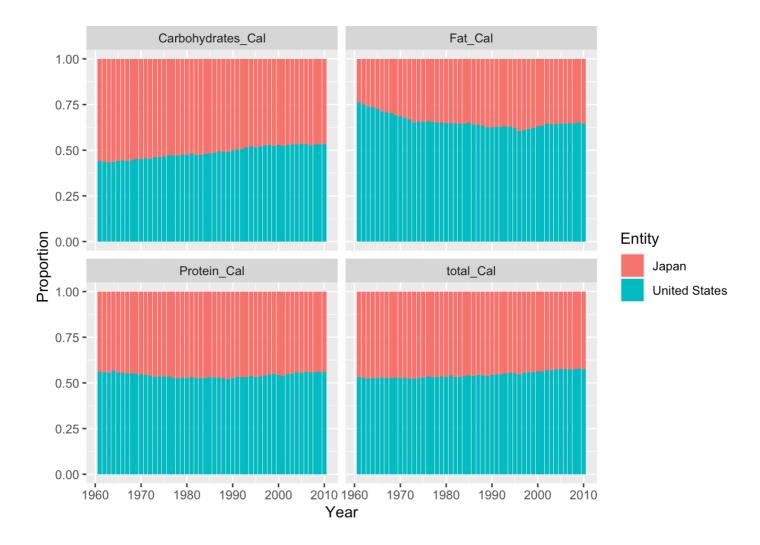
```
mydata <- mydata %>%
  mutate(total Cal = Calories.from.animal.protein..FAO..2017..
         +Calories.from.plant.protein..FAO..2017..
         +Calories.from.fat..FAO..2017..
         +Calories.from.carbohydrates..FAO..2017..,
         Protein Cal=Calories.from.animal.protein..FAO..2017..
         +Calories.from.plant.protein..FAO..2017..,
         `Protein(%)`=percent((
           Calories.from.animal.protein..FAO..2017..
           +Calories.from.plant.protein..FAO..2017..)/total_Cal,
           accuracy = 4),
         `Fat(%)`=percent(
           Calories.from.fat..FAO..2017../total_Cal
           ,accuracy = 4),
         `Carbohydrates(%)`=percent(
           Calories.from.carbohydrates..FAO..2017../total Cal,
           accuracy = 4))%>%
  rename(Fat_Cal=Calories.from.fat..FAO..2017..,
         Carbohydrates_Cal=Calories.from.carbohydrates..FAO..2017..,
         Animal Protein Cal=Calories.from.animal.protein..FAO..2017..,
         Plant Protein Cal=Calories.from.plant.protein..FAO..2017..)
mydata=mydata%>%
  filter(Year<=2010)%>%
  filter(Year>=1961)
US_Data <- mydata %>% filter(Entity =="United States")
Japan_Data <- mydata %>% filter(Entity == "Japan")
mydata <- US Data %>% rbind(Japan Data)
mydata long <- mydata %>%
  pivot_longer(cols=c(total_Cal,Protein_Cal, Fat_Cal,Carbohydrates_Cal ),
               names to = "impact variable", values to = "measure")
```

4.1 Research Question1

What is the difference in the proportions of total Calories and the three major nutrients (protein, fat, carbohydrate) from 1970 in the American and Japanese diets?

Table 4.1: Proportions of the Nutrients
Comparision

YearUnited	States	lapan
2010	3650	2685
2009	3645	2675
2008	3700	2734
2007	3757	2817
2006	3783	2778
2005	3828	2829
2004	3809	2842
2003	3777	2842
2002	3783	2853
2001	3707	2889
2000	3755	2899
1999	3673	2897
1998	3658	2895
1997	3648	2938
1996	3587	2963
1995	3580	2920
1994	3665	2932
1993	3605	
1992	3559	2943
1991	3522	2934
1990	3493	2948
1989	3433	2969
1988	3458	
1987	3450	
1986	3352	
1985	3380	
1984	3275	2827
1983	3230	2829
1982	3191	2813
1981	3218	
1980	3178	
1979	3214	
1978	3155	
1977	3135	
1976	3163	
1975	3033	
1974	3031	
1973	3040	
1972	3062	
1971		2728
1970	3029	2737



4.2 Data Explanation

The figure and table show that from 1961 to 2010, the share of per capita calorie intake in the United States and Japan did not change much, with the United States consistently having slightly higher calorie intake than Japan. From the perspective of the proportion of the three major nutrients of protein, fat and carbohydrates, the intake of fat in the American people's diet is much higher than that of the Japanese, and the intake of carbohydrates in the daily diet of the Japanese is higher than that of the United States. people. For protein intake, Americans and Japanese intakes are not much different.

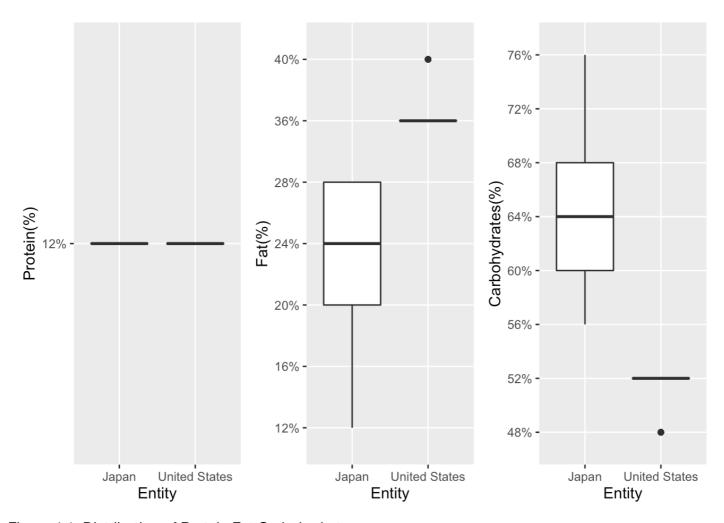


Figure 4.1: Distribution of Protein, Fat, Carbohydrates

4.3 Data Explanation

Figure 4.1 shows that in terms of diet, the difference in the proportion of protein calories consumed in Japan and the United States is not large, and the values are both around 12%. The proportion of fat and carbohydrates in the calorie intake in Japan and the United States is quite different. The proportion of fat in the Japanese diet is mostly between 20% and 28%, while the proportion of fat in the American diet is between 36% and 38%. between. Carbohydrates, on the other hand, are mostly between 60% and 68% carbohydrates in the Japanese diet, compared to 50% to 52% in the American diet.

4.4 Research Question2

What is the difference between the time trends of TotalCalories and Calories of Protein, Fat, Carbohydrates in the two countries?

Table analysis of both countries

```
summary_US <- US_Data %>%
  dplyr::select(total_Cal,Protein_Cal,Fat_Cal,Carbohydrates_Cal) %>%
  summary() %>%
  knitr::kable(caption = "Calories Intake of United States") %>%
     kable_styling(latex_options = "hold_position")
summary_US
```

Table 4.2: Calories Intake of United States

total_Cal	Protein_Cal	Fat_Cal	Carbohydrates_Cal
Min. :2858	Min. :378.3	Min. : 982.2	Min. :1481
1st Qu.:3043	1st Qu.:396.0	1st Qu.:1077.5	1st Qu.:1578
Median :3366	Median :419.8	Median :1237.6	Median :1680
Mean :3354	Mean :420.8	Mean :1221.4	Mean :1711
3rd Qu.:3650	3rd Qu.:448.5	3rd Qu.:1303.3	3rd Qu.:1863
Max. :3828	Max. :461.9	Max. :1484.9	Max. :1941

```
summary_Japan <- Japan_Data %>%
  dplyr::select(total_Cal,Protein_Cal,Fat_Cal,Carbohydrates_Cal) %>%
  summary() %>%
  knitr::kable(caption = "Calories Intake of Japan") %>%
     kable_styling(latex_options = "hold_position")

summary_Japan
```

Table 4.3: Calories Intake of Japan

total_Cal	Protein_Cal	Fat_Cal	Carbohydrates_Cal
Min. :2525	Min. :296.8	Min. :310.4	Min. :1547
1st Qu.:2730	1st Qu.:339.9	1st Qu.:558.4	1st Qu.:1727
Median :2810	Median :359.5	Median :694.9	Median :1800
Mean :2800	Mean :357.2	Mean :652.0	Mean :1790
3rd Qu.:2895	3rd Qu.:380.6	3rd Qu.:791.8	3rd Qu.:1860
Max. :2969	Max. :392.9	Max. :815.5	Max. :1962

4.5 Data Explanation

It can be seen from Table4.2, that the mean Calories of United States is 3354 kcal while Table4.3 shows Japan's mean Calories is 2800. In addition, we can observe that the average carbohydrate intake of the Japanese and American diets is almost the same in terms of the average calorie intake of the three nutrients, but the fat intake of the American diet is significantly higher than that of the Japanese diet.

```
plot_Calories_Intake <- mydata%>%
    ggplot(section2_chile_canada, mapping = aes(
        x = Year,
        y = Protein_Cal,
        color = Entity)) +
    geom_line() +
    theme_bw() +
    xlab("Year") +
    ylab("Total Calories Intake") +
    ggtitle("Calories Intake of over the years")
plot_Calories_Intake
```

Calories Intake of over the years

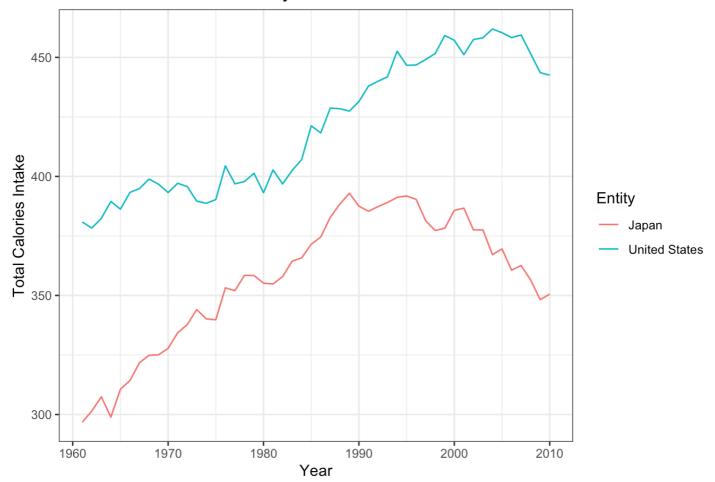


Figure 4.2: Calories Intake of over the years

4.6 Data Explanation

Figure 4.2 shows the trend of total calories intake in the United States and Japan over time. From the point of total dietary calorie intake, Figure 4.2 shows that dietary calorie intake in Japan first increased over the past 50 years and then gradually decreased after reaching a peak around 1995. In the United States, diets continued to increase until they began to decrease after 2000. The calorie intake gap between the two countries first decreased and then gradually increased.

```
plot_Fat_Calories_Intake <- mydata%>%
    ggplot(section2_chile_canada, mapping = aes(
        x = Year,
        y = Fat_Cal,
        color = Entity)) +
    geom_line() +
    theme_bw() +
    xlab("Year") +
    ylab("Fat Calories Intake") +
    ggtitle("Fat Calories Intake of over the years")
plot_Fat_Calories_Intake
```

Fat Calories Intake of over the years

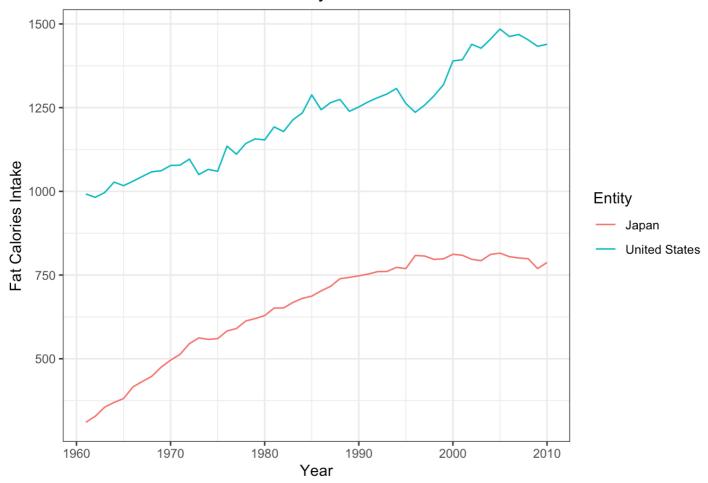


Figure 4.3: Fat Calories Intake of over the years

4.7 Data Explanation

Figure 4.3 shows the trend of fat calories intake in the United States and Japan over time. From the Figure 4.3, we can find that the intake of fat in the diet of Japan and the United States shows a trend of increasing year by year, and the gap between the two countries has changed very little in the past 50 years, and it can be seen as almost no change.

5 Analysis4

5.1 Research Question

Rice consumption vs. latitude and region, 2015

```
# Data
Assignment4_data <- read_csv("Data/rice-consumption-vs-latitude.csv")
data_tidy <- Assignment4_data %>%
   filter(Year == 2015)%>%
   rename(`Rice consumption(kg/capita/yr)` = `Rice (Milled Equivalent) - Food suppl
y quantity (kg/capita/yr)`) %>%
   rename(Latitude = `Latitude - lp_lat_abst`)

data_tidy <- data_tidy %>% drop_na(`Rice consumption(kg/capita/yr)`) %>% drop_na(Latitude)
```

```
knitr::kable(
  head(arrange(data_tidy,desc(`Rice consumption(kg/capita/yr)`)),10), caption = 'T
  op 10 countries with the highest annual per capita consumption of rice in 2015',
    booktabs = TRUE,digits = 2
) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

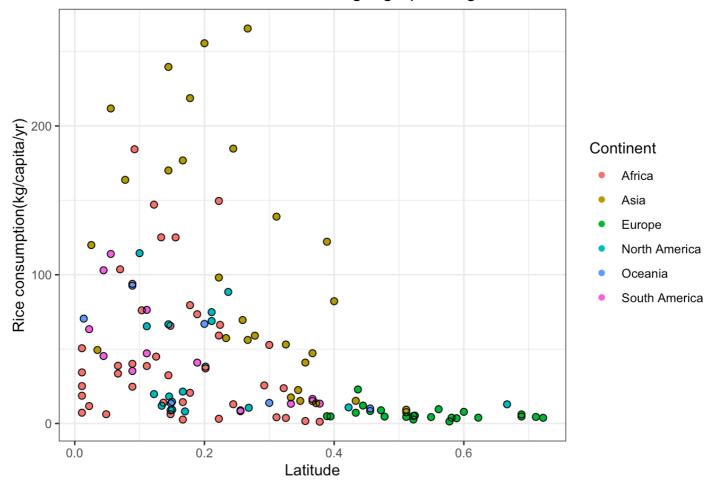
Table 5.1: Top 10 countries with the highest annual per capita consumption of rice in 2015

Entity	Code	Year	Rice consumption(kg/capita/yr)	Latitude	Continent
Bangladesh	BGD	2015	265.55	0.27	Asia
Laos	LAO	2015	255.64	0.20	Asia
Cambodia	KHM	2015	239.70	0.14	Asia
Vietnam	VNM	2015	218.73	0.18	Asia
Indonesia	IDN	2015	211.79	0.06	Asia
Myanmar	MMR	2015	184.80	0.24	Asia

Sierra Leone	SLE	2015	184.36	0.09	Africa
Thailand	THA	2015	176.85	0.17	Asia
Philippines	PHL	2015	170.10	0.14	Asia
Sri Lanka	LKA	2015	163.80	0.08	Asia

```
ggplot(data = data_tidy,
aes(x = Latitude,
    y = `Rice consumption(kg/capita/yr)`)) +
geom_point(aes(colour = Continent)) +
geom_point(data = data_tidy,
size = 2,
shape = 1)+
theme_bw()+
ggtitle("Distribution of countries in different geographic regions in terms of L
atitude")
```

Distribution of countries in different geographic regions in terms of Latitude



5.2 Data Explanation

Table 1 ranks all the Annual per capita consumption of rice in different countries in 2015 in descending order, while Figure 1 plots the distribution by Latitude and the geographical region to which the country belongs. According to Table 1, the top 5 countries with the largest rice consumption are Bangladesh, Laos, Cambodia, Vietnam,In combination with Figure 1, it is easy to notice the phenomenon that the countries with higher Annual per capita consumption of rice are mainly in the Latitude between and 0.4. Also, when looking at the color of the points, it can be seen that the points representing higher rice consumption represent the map areas of Africa and Asia.

6 Conclusion

According to the Australian government's dietary guidelines, these unhealthy diets have led to many Australian adults and about a quarter of children being overweight or obese, so it's time to make changes for the sake of our health Grech, Rangan, and Allman-Farinelli (2018).

All in all, with the development of society and the progress of economy, how to maintain a healthy eating habit has become an increasingly important issue Unger et al. (1992). In addition to paying attention to the total calorie intake of the diet, people also need to pay attention to the energy supply ratio of the three major nutrients, protein, fat and carbohydrates. Reasonable arrangement of the proportion of nutrients can help us maintain a healthier body and prolong our energy consumption. longevity and reduce the incidence of disease Lands et al. (1990).

Reference

- Grech, Amanda, Anna Rangan, and Margaret Allman-Farinelli. 2018. "Macronutrient Composition of the Australian Population's Diet; Trends from Three National Nutrition Surveys 1983, 1995 and 2012." *Nutrients* 10 (8): 1045.
- Lands, W. E., T. Hamazaki, K. Yamazaki, H. Okuyama, K. Sakai, Y. Goto, and V. S. Hubbard. 1990. "Changing Dietary Patterns." *American Journal of Clinical Nutrition* 51 (6): 991–93.
- Stoll-Kleemann, Susanne, and Tim O'Riordan. 2015. "The Sustainability Challenges of Our Meat and Dairy Diets." *Environment: Science and Policy for Sustainable Development* 57 (3): 34–48.
- Unger, R., M. Dekleermaeker, S. S. Gidding, and K. K. Christoffel. 1992. "Calories Count. Improved Weight Gain with Dietary Intervention in Congenital Heart Disease." *American Journal of Diseases of Children* 146 (9): 1078.
- Yang, Shengping, and Gilbert Berdine. 2015. "Model Selection and Model over-Fitting." *The Southwest Respiratory and Critical Care Chronicles* 3 (12): 52–55.