Assignment1

Data:

The data set is from Kaggle containing the McDonalds menu with nutrition information and some notes are available using R and python with plotly. Some demos in (https://www.kaggle.com/limitpointinf0/mc-d-nutrition-eda) and just keep a note for the talk https://github.com/jrauser/writing/blob/master/how_humans_see_data/hhsd_notes.Rmd

I eat a lot of fast food and McDonalds is one of my most usually visit restaurant after midnight. So I am very interested in how I gain so much weights, so this data set is fairly interesting to me.
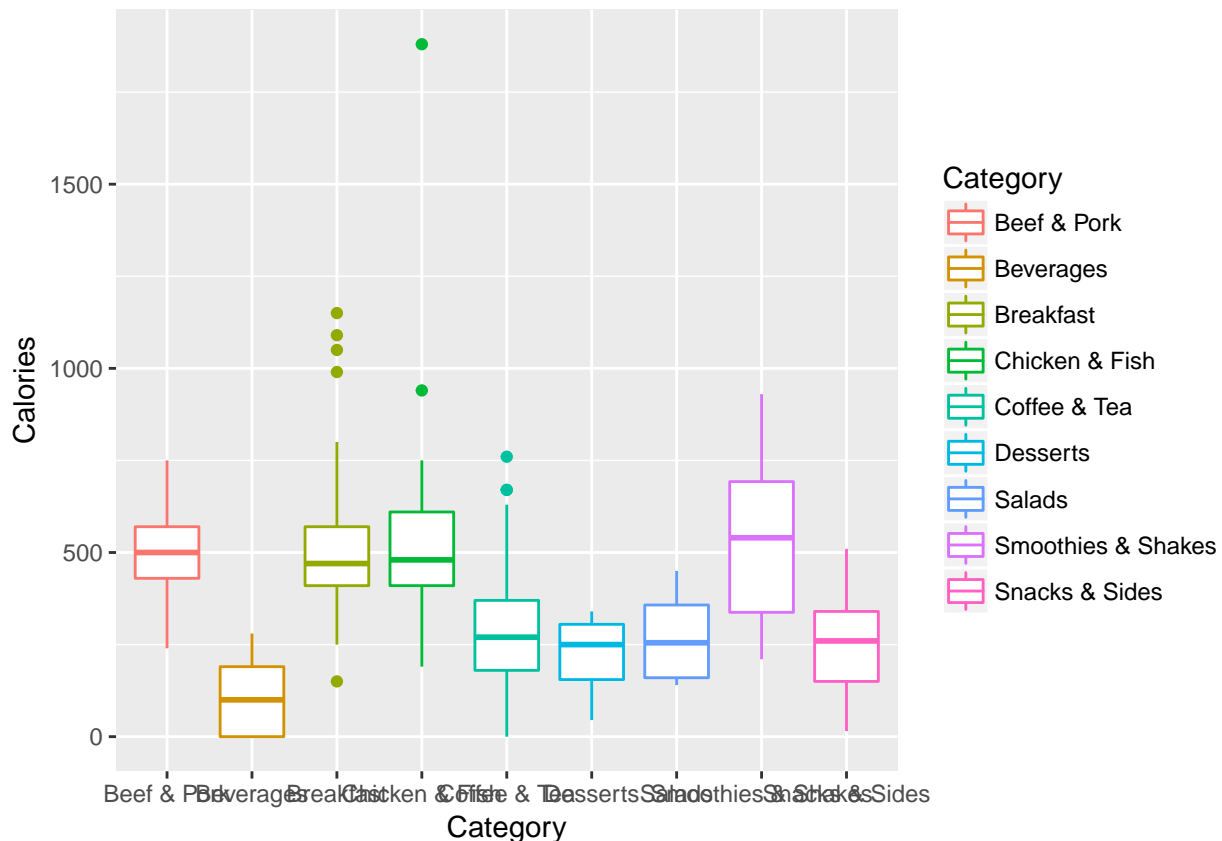
Graphs:

```
library(tidyverse)
library(ggplot2)
aa<-read.csv('menu.csv')
bb<-as_data_frame(aa)
```

Here, I have a question, why is that if I use read_csv('menu.csv'), this considers some categorigal data as char, so I have to use.function and change into dataframe. Can I have a way fixing it?

Graph 1:

```
gg1<-ggplot(bb, aes(x=Category ,y=Calories, colour = Category)) +geom_boxplot()
gg1
```

This is very interesting for me that beverage has the least calories, I always thought the pop makes me fat.

```
result <- bb %>%
  select(Category,Item,Serving.Size,Calories) %>% group_by(Category) %>% top_n(3)# %>% summarise(Catego
```

```
## Selecting by Calories
```

```
data.frame(result[,c('Item','Calories')])
```

```
##                                                         Item Calories
## 1                Big Breakfast with Hotcakes (Regular Biscuit)     1090
## 2                  Big Breakfast with Hotcakes (Large Biscuit)     1150
## 3   Big Breakfast with Hotcakes and Egg Whites (Large Biscuit)     1050
## 4                   Quarter Pounder with Bacon Habanero Ranch      610
## 5                        Double Quarter Pounder with Cheese      750
## 6                                   Bacon Clubhouse Burger      720
## 7                  Bacon Clubhouse Crispy Chicken Sandwich      750
## 8                            Chicken McNuggets (20 piece)      940
## 9                            Chicken McNuggets (40 piece)     1880
## 10       Premium Bacon Ranch Salad with Crispy Chicken      380
## 11         Premium Southwest Salad with Crispy Chicken      450
## 12         Premium Southwest Salad with Grilled Chicken      290
## 13            Chipotle BBQ Snack Wrap (Crispy Chicken)      340
## 14                 Ranch Snack Wrap (Crispy Chicken)      360
## 15                              Medium French Fries      340
## 16                               Large French Fries      510
## 17                                 Hot Fudge Sundae      330
## 18                               Hot Caramel Sundae      340
## 19                                Strawberry Sundae      280
## 20                        Coca-Cola Classic (Large)      280
## 21                                   Sprite (Large)      280
## 22                   Minute Maid Orange Juice (Large)      280
## 23                             Frappé Mocha (Large)      670
## 24                           Frappé Caramel (Large)      670
## 25                  Frappé Chocolate Chip (Large)      760
## 26                         Strawberry Shake (Large)      850
## 27                          Chocolate Shake (Large)      850
## 28               McFlurry with M&M's Candies (Medium)      930
```

We can see from the select top 3 highest calories food from each Category, which is related with the size.

Graph 2:

```
(gg2<-ggplot(bb,aes(x=Calories,y=Saturated.Fat,colour=Category))
+geom_point()+facet_wrap(~Category)+ geom_smooth(method='loess')+theme_minimal())
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 138.45
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 151.55
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 6400
```
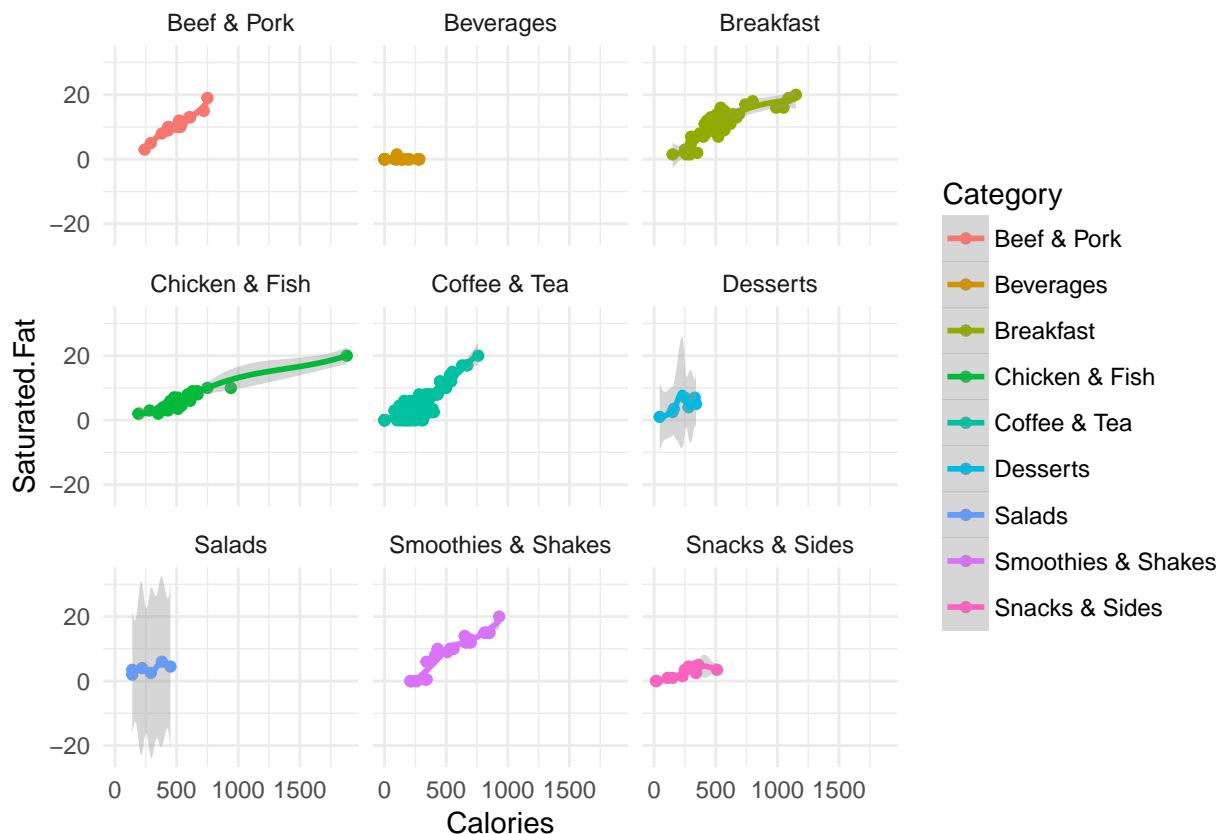
```
## Warning in predLoess(object$y, object$x, newx = if
```

```
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used
## at 138.45

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 151.55

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal
## condition number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other
## near singularities as well. 6400
```



```
#scale_colour_hue(h = c(270, 360))+
gg2
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 138.45

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 151.55

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0
```
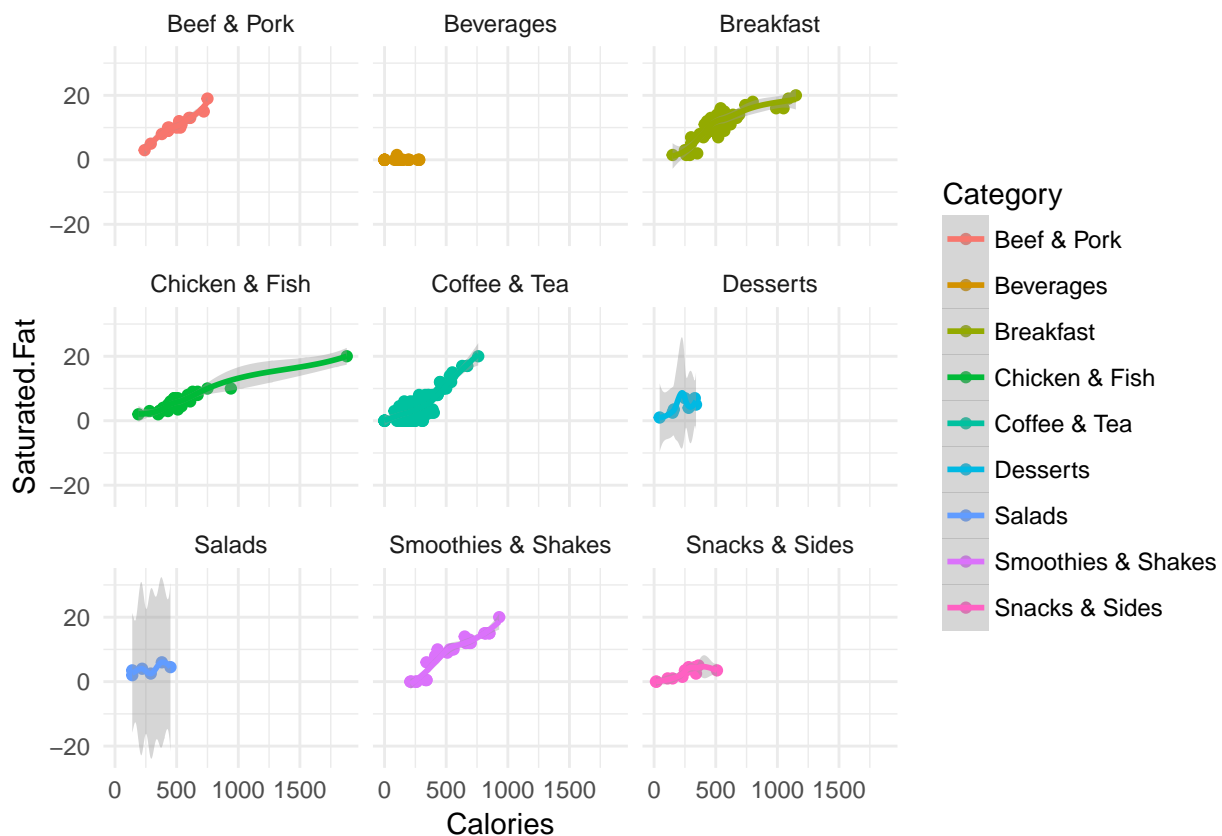
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 6400

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used
## at 138.45

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 151.55

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal
## condition number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other
## near singularities as well. 6400
```


From the second graph, we can tell that bad fat has postive linear correlation with calories, for beverages, salads and snacks & Sides, the slopes are rather soft, which means that even though they have high calories, they do not contribute much to my fatness.