

Untitled

Ahn Chiung

2019 10 4

0.1 2014-17374

0.1.1 Problem 1.

Problem 1.1.

```
library(openintro)
```

```
## Please visit openintro.org for free statistics materials
```

```
##
```

```
## Attaching package: 'openintro'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
##     diamonds
```

```
## The following objects are masked from 'package:datasets':
```

```
##
```

```
##     cars, trees
```

```
data(starbucks)
```

```
?starbucks
```

```
## starting httpd help server ...
```

```
## done
```

```
starbucks
```

```
##
```

	item	calories	fat
## 1	8-Grain Roll	350	8.0
## 2	Apple Bran Muffin	350	9.0
## 3	Apple Fritter	420	20.0
## 4	Banana Nut Loaf	490	19.0
## 5	Birthday Cake Mini Doughnut	130	6.0
## 6	Blueberry Oat Bar	370	14.0
## 7	Blueberry Scone	460	22.0
## 8	Bountiful Blueberry Muffin	370	14.0
## 9	Butter Croissant	310	18.0
## 10	Cheese Danish	420	25.0

## 11	Chocolate Chunk Cookie	380 17.0
## 12	Chocolate Cinnamon Bread	320 12.0
## 13	Chocolate Croissant	300 17.0
## 14	Chocolate Old-Fashioned Doughnut	420 21.0
## 15	Chonga Bagel	310 5.0
## 16	Cinnamon Chip Scone	480 18.0
## 17	Cranberry Orange Scone	490 18.0
## 18	Double Chocolate Brownie	410 24.0
## 19	Double Fudge Mini Doughnut	130 7.0
## 20	Everything with Cheese Bagel	280 2.0
## 21	Ginger Molasses Cookie	360 12.0
## 22	Iced Lemon Pound Cake	490 23.0
## 23	Mallorca Sweet Bread	420 25.0
## 24	Maple Oat Pecan Scone	440 18.0
## 25	Marble Pound Cake	350 13.0
## 26	Marshmallow Dream Bar	210 4.0
## 27	Morning Bun	350 16.0
## 28	Multigrain Bagel	300 3.0
## 29	Old-Fashioned Glazed Doughnut	420 21.0
## 30	Outrageous Oatmeal Cookie	370 14.0
## 31	Petite Vanilla Bean Scone	140 5.0
## 32	Plain Bagel	280 1.0
## 33	Pumpkin Bread	390 14.0
## 34	Pumpkin Scone	480 17.0
## 35	Raspberry Scone	480 25.0
## 36	Raspberry Swirl Pound Cake	430 16.0
## 37	Reduced-Fat Banana Chocolate Chip Coffee Cake	400 8.0
## 38	Reduced-Fat Cinnamon Swirl Coffee Cake	340 9.0
## 39	Reduced-Fat Very Berry Coffee Cake	350 10.0
## 40	Starbucks Classic Coffee Cake	440 19.0
## 41	Zucchini Walnut Muffin	490 28.0
## 42	Cheese & Fruit	480 28.0
## 43	Chicken & Hummus	270 8.0
## 44	Chicken Lettuce Wraps	360 19.0
## 45	Chipotle Chicken Wraps	380 15.0
## 46	Protein	380 19.0
## 47	Salumi & Cheese	420 26.0
## 48	Sesame Noodles	350 11.0
## 49	Tuna Salad	380 21.0
## 50	Apple Pie	180 7.0
## 51	Birthday Cake Pop	170 9.0

## 52				Brown Sugar Walnut Tart	190 12.0
## 53				Cherry Pie	170 7.0
## 54				Chocolate Creme Whoopie Pie	190 11.0
## 55				Chocolate Hazelnut Tart	180 10.0
## 56				Raspberry Truffle Cake Pop	160 8.0
## 57				Red Velvet Whoopie Pie	190 11.0
## 58				Tiramisu Cake Pop	170 9.0
## 59				Bacon & Gouda Artisan Breakfast Sandwich	350 18.0
## 60				Chicken Sausage Breakfast Wrap	300 10.0
## 61				Ham & Cheddar Artisan Breakfast Sandwich	350 16.0
## 62				Sausage & Cheddar Classic Breakfast Sandwich	500 28.0
## 63				Spinach & Feta Breakfast Wrap	290 10.0
## 64				Starbucks Perfect Oatmeal	140 2.5
## 65				Turkey Bacon & White Cheddar Classic Breakfast Sandwich	320 7.0
## 66				Veggie & Monterey Jack Artisan Breakfast Sandwich	350 18.0
## 67				Deluxe Fruit Blend	80 0.0
## 68				Chicken Santa Fe Panini	400 11.0
## 69				Egg Salad Sandwich	460 27.0
## 70				Ham & Swiss Panini	360 9.0
## 71				Roasted Tomato & Mozzarella Panini	390 18.0
## 72				Roasted Vegetable Panini	350 12.0
## 73				Tarragon Chicken Salad Sandwich	420 13.0
## 74				Turkey & Swiss Sandwich	390 13.0
## 75				Greek Yogurt Honey Parfait	300 12.0
## 76				Peach Raspberry Yogurt Parfait	300 4.0
## 77				Strawberry & Blueberry Yogurt Parfait	300 3.5
##	carb	fiber	protein	type	
## 1	67	5	10	bakery	
## 2	64	7	6	bakery	
## 3	59	0	5	bakery	
## 4	75	4	7	bakery	
## 5	17	0	0	bakery	
## 6	47	5	6	bakery	
## 7	61	2	7	bakery	
## 8	55	0	6	bakery	
## 9	32	0	5	bakery	
## 10	39	0	7	bakery	
## 11	51	2	4	bakery	
## 12	53	3	6	bakery	
## 13	34	2	5	bakery	
## 14	57	2	5	bakery	

## 15	52	3	12	bakery
## 16	70	3	7	bakery
## 17	73	2	8	bakery
## 18	46	3	6	bakery
## 19	16	0	0	bakery
## 20	56	2	10	bakery
## 21	58	0	3	bakery
## 22	67	0	5	bakery
## 23	42	0	7	bakery
## 24	59	3	8	bakery
## 25	54	0	6	bakery
## 26	43	0	0	bakery
## 27	45	2	6	bakery
## 28	60	6	15	bakery
## 29	57	0	4	bakery
## 30	56	3	5	bakery
## 31	21	0	0	bakery
## 32	59	2	9	bakery
## 33	61	2	6	bakery
## 34	78	2	6	bakery
## 35	59	3	8	bakery
## 36	69	0	4	bakery
## 37	80	4	5	bakery
## 38	62	2	4	bakery
## 39	59	4	7	bakery
## 40	63	0	6	bakery
## 41	52	2	7	bakery
## 42	39	6	18	bistro box
## 43	29	6	16	bistro box
## 44	32	4	17	bistro box
## 45	35	6	26	bistro box
## 46	37	5	13	bistro box
## 47	22	3	25	bistro box
## 48	50	6	15	bistro box
## 49	25	5	23	bistro box
## 50	27	0	2	petite
## 51	22	0	0	petite
## 52	24	0	2	petite
## 53	24	0	2	petite
## 54	23	0	0	petite
## 55	23	0	2	petite

```
## 56 24 0 2 petite
## 57 21 0 0 petite
## 58 22 0 0 petite
## 59 30 0 17 hot breakfast
## 60 33 5 14 hot breakfast
## 61 31 0 20 hot breakfast
## 62 41 0 19 hot breakfast
## 63 33 6 19 hot breakfast
## 64 25 4 5 hot breakfast
## 65 43 3 18 hot breakfast
## 66 30 0 17 hot breakfast
## 67 20 2 0 salad
## 68 47 2 26 sandwich
## 69 37 5 22 sandwich
## 70 43 2 28 sandwich
## 71 44 3 15 sandwich
## 72 48 4 13 sandwich
## 73 46 6 32 sandwich
## 74 36 2 34 sandwich
## 75 44 0 8 parfait
## 76 57 3 10 parfait
## 77 60 3 7 parfait
```

결측치가 있는지 확인해본 결과, 다음과 같이 결측치는 존재하지않았다.

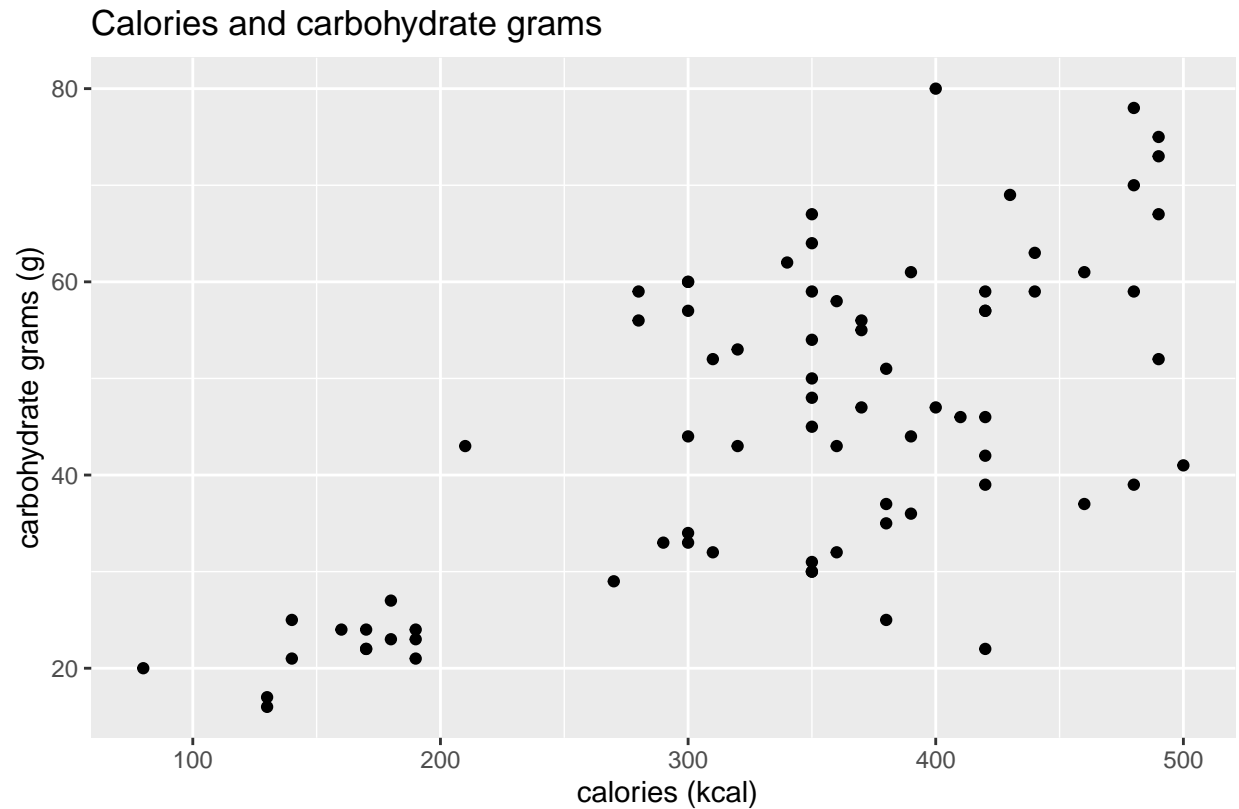
```
sum(is.na(starbucks))
```

```
## [1] 0
```

help에서 찾은 Source 인 <http://www.starbucks.com/menu/nutrition> 들어가 찾아본 결과, 사이트가 만료되었고, 현재 메뉴들은 명칭이 바뀌어 있다. (<https://www.starbucks.com/menu>) 현재 표기법을 기반으로, 단위를 유추해보면, 칼로리는 kcal, 지방은 g, 탄수화물은 g, 식이섬유는 g, 단백질은 g 이다.

scatter plot을 그려보면 다음과 같다.

```
ggplot(starbucks, mapping = aes(x = calories, y = carb)) +
  geom_point(mapping = aes(x = calories, y = carb))+
  labs(
    title = "Calories and carbohydrate grams",
    caption = "Data on March 10, 2011",
    x = " calories (kcal)",
    y = "carbohydrate grams (g)"
  )
```



```
?geom_smooth
```

일단 칼로리가 250kcal이하인 것이 적다. 한편, 탄수화물은 데이터에서 20g이하 70g이상을 제외하고는 고르게 분포하고 있는 것 같다.

또한, 칼로가 250kcal이하에서는 대부분 칼로리가 200이하면서 탄수화물이 30g으로 적게 함유되어 있지만, 예외적으로 군집에서 떨어져 보이는 점이 보인다. 칼로리가 200에서 250사이에 유일한 점인데, 탄수화물이 40에서 50사이이다.

칼로리가 250초과인 점들은 탄수화물이 20이상부터 80까지 넓게 퍼져있다.

한편, 전체적으로는 칼로리와 탄수화물 함량이 양의 상관관계로 보인다.

Problem 1.2.

```
attach(starbucks)
fit1<-lm(carb~calories)
summary(fit1)
```

```
##
## Call:
## lm(formula = carb ~ calories)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.477  -7.476  -1.029   10.127   28.644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.94356    4.74600   1.884   0.0634 .
## calories      0.10603    0.01338   7.923 1.67e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.29 on 75 degrees of freedom
## Multiple R-squared:  0.4556, Adjusted R-squared:  0.4484
## F-statistic: 62.77 on 1 and 75 DF,  p-value: 1.673e-11
```

Problem 1.3.

```
fc <- fit1$coefficients
fc[1]
```

```
## (Intercept)
##      8.94356
```

```
fc[-1]
```

```
## calories
## 0.1060309
```

```
names(fc[-1])
```

```
## [1] "calories"
```

```
mathexpression <- paste("carbohydrate grams =", paste(round(fc[1],2), paste(round(fc[-1],2), names(fc[-1]))))
mathexpression
```

```
## [1] "carbohydrate grams = 8.94 + 0.11 * calories + e"
```

먼저 intercept는 calories가 0kcal일때 추정되는 탄수화물 함량으로서 8.94g이다. 또한 slope는 calories가 1kcal 증가할 때, 평균적으로 증가하는 탄수화물의 함량으로 0.11g이다.

Problem 1.4.

```
summary(fit1)["r.squared"]
```

```
## $r.squared
## [1] 0.4556237
```

```
sqrt(0.4556237)
```

```
## [1] 0.674999
```

위와 같이 이 모델의 Multiple R-squared는 0.4556이다. 해석은 전체 변동(SST)중 모델이 설명하는 변동(SSR)의 비율이 약 45.56%라고 할 수 있다는 뜻이다. 따라서 모델을 선택할 때 잔차를 잘 설명하는 모델로서 Multiple R-squared 값이 큰 것으로 선택할 수 있다. 더하여서, 피어슨의 상관계수의 제곱이 Multiple R-squared 이므로 선형 상관관계의 강도를 나타는 것으로도 해석될 수 있다. 이때 45.56%정도만 설명하므로 50%를 기준으로 한다면, 50%이하로 충분한 모델이라고 볼 수 없다.

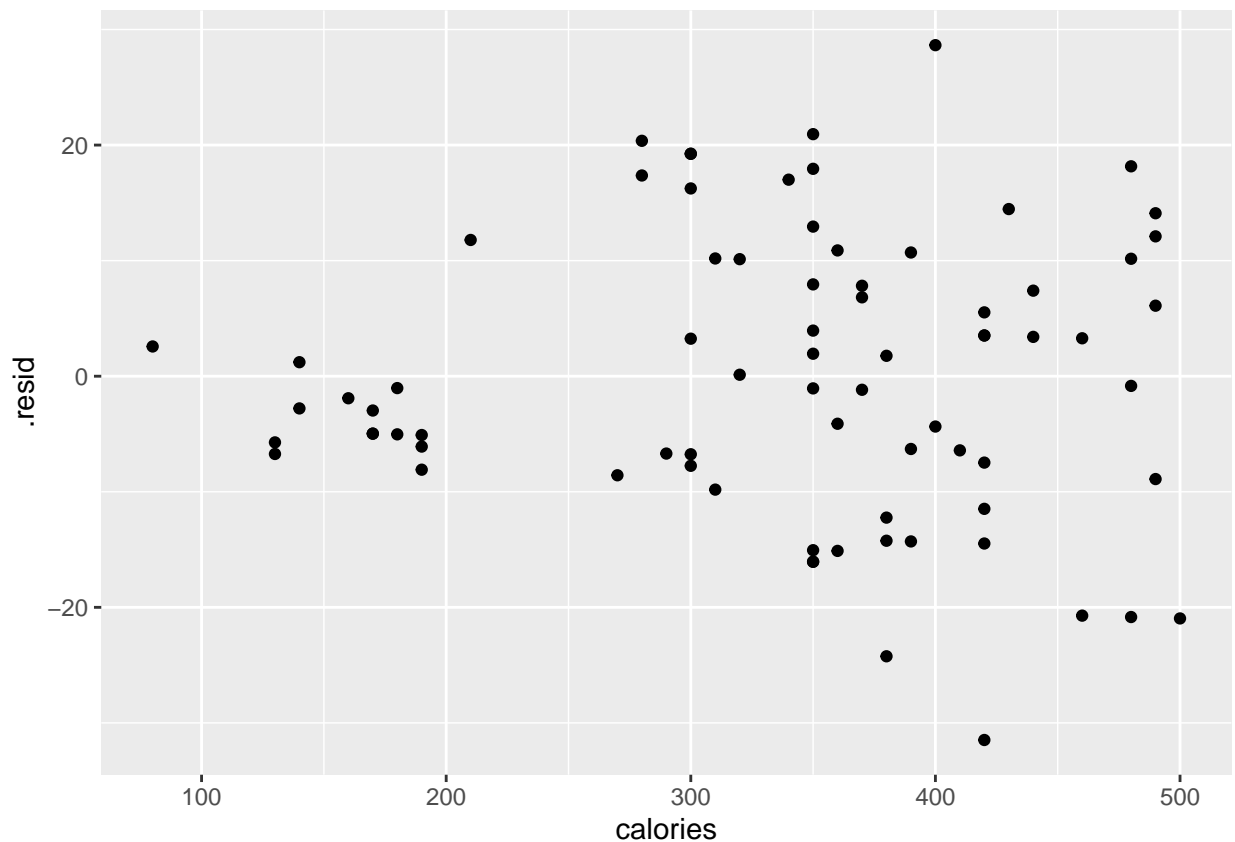
```
?fortify
```

```
inf1<-fortify(fit1, starbucks)
```

```
names(inf1)
```

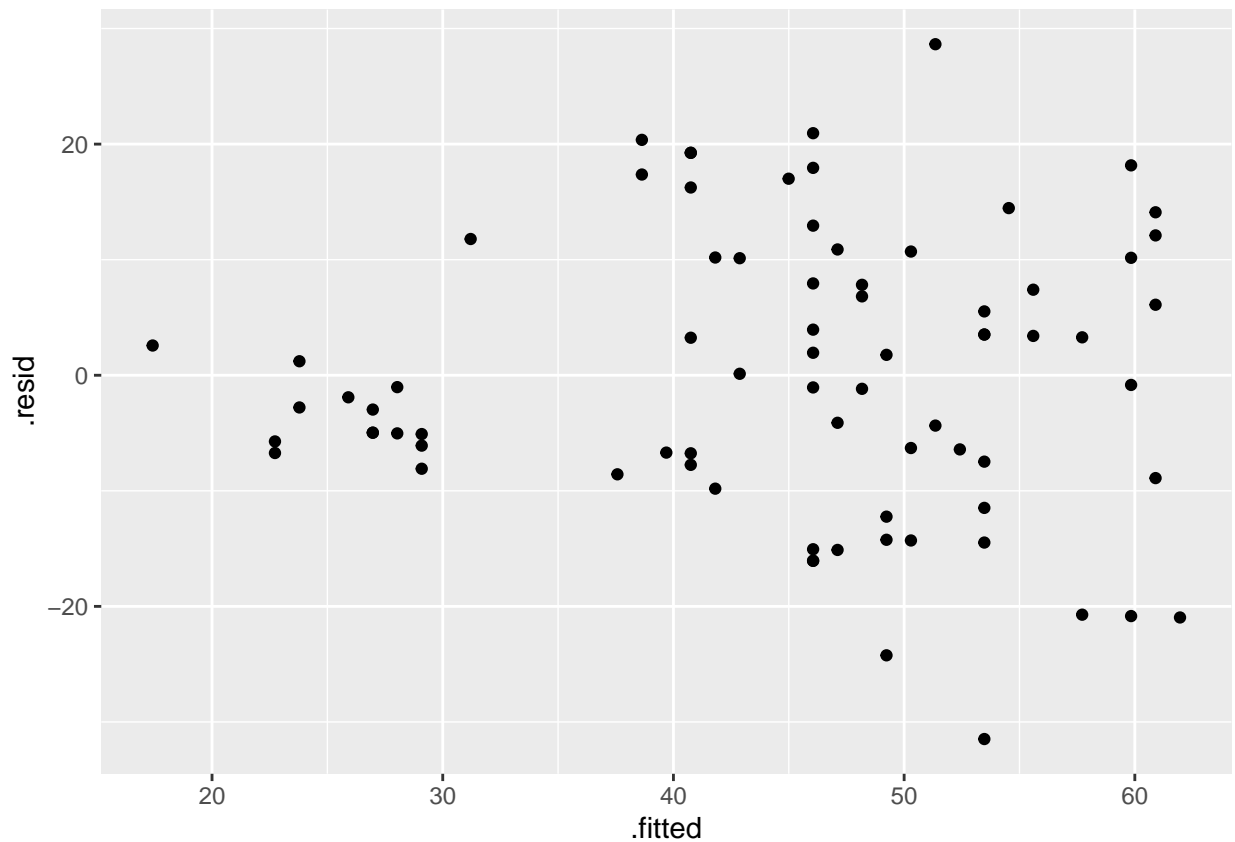
```
## [1] "item"      "calories"  "fat"       "carb"      "fiber"
## [6] "protein"   "type"      ".hat"       ".sigma"     ".cooks"
## [11] ".fitted"   ".resid"    ".stdresid"
```

```
ggplot(inf1, aes(x = calories, y = .resid))+
  geom_point()
```



x axis를 calories로 그린 residual plot 이다.


```
ggplot(inf1, aes(x = .fitted, y = .resid))+
  geom_point()
```



x axis를 fitted value로 그린 residual plot 이다.

위 두 residual plot을 통해서 알 수 있는 사실은 0을 기준으로 분산이 전반적으로 대칭적인 모습을 하고 있지만, calories 값이 커질 수록 residual의 산포가 넓어지는 것을 알 수 있다.

따라서 등분산성을 가정할 수 없고, weighted least square 방법을 이용하여 추정해야한다.

이때 계량경제학시간에서 이용한 weight를 구하는 방법을 이용한다.

$\log(e^2)$ 값을 독립변수들로 ols로 추정하고, 나온 추정량을 다시 $\exp(\text{estimated } \log(e^2))$ 로 추정 분산을 구해서 weight 값으로 $1/\exp(\text{estimated } \log(e^2))$ 를 대입한다.

코드와 결과는 다음과 같다.

```
lerror1<-log((fit1$residuals)^2)
result1_f<-lm(lerror1~calories)
pred1<-exp(result1_f$fitted.values)
mod1<-1/pred1
result1_2<-lm(carb~calories, weights=mod1)
summary(result1_2)
```

```
##
## Call:
## lm(formula = carb ~ calories, weights = mod1)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6853 -1.0636 -0.1606  1.0407  3.3751
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.47500    3.19069   2.029   0.046 *
## calories     0.11402    0.01072  10.637 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.625 on 75 degrees of freedom
## Multiple R-squared:  0.6014, Adjusted R-squared:  0.5961
## F-statistic: 113.2 on 1 and 75 DF,  p-value: < 2.2e-16
```

Multiple R-squared값이 0.6014로 증가한 것을 살펴볼 수 있다.

```
detach(starbucks)
```

0.1.2 Problem 2.

```
absenteeism<-read.csv("absenteeism.csv")
head(absenteeism)
```

```
##   X eth sex age lrn days
## 1 1   A   M  FO  SL    2
## 2 2   A   M  FO  SL   11
## 3 3   A   M  FO  SL   14
## 4 4   A   M  FO  AL    5
## 5 5   A   M  FO  AL    5
## 6 6   A   M  FO  AL   13
```

이후 package를 찾는 끝에 MASS라는 패키지에 존재함을 알 수 있었다.

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
## The following objects are masked from 'package:openintro':
##
```

```
##      housing, mammals

## The following object is masked from 'package:dplyr':
##
##      select
```

```
data(quine)
quine
```

```
##      Eth Sex Age Lrn Days
## 1      A  M  F0  SL    2
## 2      A  M  F0  SL   11
## 3      A  M  F0  SL   14
## 4      A  M  F0  AL    5
## 5      A  M  F0  AL    5
## 6      A  M  F0  AL   13
## 7      A  M  F0  AL   20
## 8      A  M  F0  AL   22
## 9      A  M  F1  SL    6
## 10     A  M  F1  SL    6
## 11     A  M  F1  SL   15
## 12     A  M  F1  AL    7
## 13     A  M  F1  AL   14
## 14     A  M  F2  SL    6
## 15     A  M  F2  SL   32
## 16     A  M  F2  SL   53
## 17     A  M  F2  SL   57
## 18     A  M  F2  AL   14
## 19     A  M  F2  AL   16
## 20     A  M  F2  AL   16
## 21     A  M  F2  AL   17
## 22     A  M  F2  AL   40
## 23     A  M  F2  AL   43
## 24     A  M  F2  AL   46
## 25     A  M  F3  AL    8
## 26     A  M  F3  AL   23
## 27     A  M  F3  AL   23
## 28     A  M  F3  AL   28
## 29     A  M  F3  AL   34
## 30     A  M  F3  AL   36
## 31     A  M  F3  AL   38
## 32     A  F  F0  SL    3
```

## 33	A	F	F0	AL	5
## 34	A	F	F0	AL	11
## 35	A	F	F0	AL	24
## 36	A	F	F0	AL	45
## 37	A	F	F1	SL	5
## 38	A	F	F1	SL	6
## 39	A	F	F1	SL	6
## 40	A	F	F1	SL	9
## 41	A	F	F1	SL	13
## 42	A	F	F1	SL	23
## 43	A	F	F1	SL	25
## 44	A	F	F1	SL	32
## 45	A	F	F1	SL	53
## 46	A	F	F1	SL	54
## 47	A	F	F1	AL	5
## 48	A	F	F1	AL	5
## 49	A	F	F1	AL	11
## 50	A	F	F1	AL	17
## 51	A	F	F1	AL	19
## 52	A	F	F2	SL	8
## 53	A	F	F2	SL	13
## 54	A	F	F2	SL	14
## 55	A	F	F2	SL	20
## 56	A	F	F2	SL	47
## 57	A	F	F2	SL	48
## 58	A	F	F2	SL	60
## 59	A	F	F2	SL	81
## 60	A	F	F2	AL	2
## 61	A	F	F3	AL	0
## 62	A	F	F3	AL	2
## 63	A	F	F3	AL	3
## 64	A	F	F3	AL	5
## 65	A	F	F3	AL	10
## 66	A	F	F3	AL	14
## 67	A	F	F3	AL	21
## 68	A	F	F3	AL	36
## 69	A	F	F3	AL	40
## 70	N	M	F0	SL	6
## 71	N	M	F0	SL	17
## 72	N	M	F0	SL	67
## 73	N	M	F0	AL	0

## 74	N	M	F0	AL	0
## 75	N	M	F0	AL	2
## 76	N	M	F0	AL	7
## 77	N	M	F0	AL	11
## 78	N	M	F0	AL	12
## 79	N	M	F1	SL	0
## 80	N	M	F1	SL	0
## 81	N	M	F1	SL	5
## 82	N	M	F1	SL	5
## 83	N	M	F1	SL	5
## 84	N	M	F1	SL	11
## 85	N	M	F1	SL	17
## 86	N	M	F1	AL	3
## 87	N	M	F1	AL	4
## 88	N	M	F2	SL	22
## 89	N	M	F2	SL	30
## 90	N	M	F2	SL	36
## 91	N	M	F2	AL	8
## 92	N	M	F2	AL	0
## 93	N	M	F2	AL	1
## 94	N	M	F2	AL	5
## 95	N	M	F2	AL	7
## 96	N	M	F2	AL	16
## 97	N	M	F2	AL	27
## 98	N	M	F3	AL	0
## 99	N	M	F3	AL	30
## 100	N	M	F3	AL	10
## 101	N	M	F3	AL	14
## 102	N	M	F3	AL	27
## 103	N	M	F3	AL	41
## 104	N	M	F3	AL	69
## 105	N	F	F0	SL	25
## 106	N	F	F0	AL	10
## 107	N	F	F0	AL	11
## 108	N	F	F0	AL	20
## 109	N	F	F0	AL	33
## 110	N	F	F1	SL	5
## 111	N	F	F1	SL	7
## 112	N	F	F1	SL	0
## 113	N	F	F1	SL	1
## 114	N	F	F1	SL	5

```
## 115 N F F1 SL 5
## 116 N F F1 SL 5
## 117 N F F1 SL 5
## 118 N F F1 SL 7
## 119 N F F1 SL 11
## 120 N F F1 SL 15
## 121 N F F1 AL 5
## 122 N F F1 AL 14
## 123 N F F1 AL 6
## 124 N F F1 AL 6
## 125 N F F1 AL 7
## 126 N F F1 AL 28
## 127 N F F2 SL 0
## 128 N F F2 SL 5
## 129 N F F2 SL 14
## 130 N F F2 SL 2
## 131 N F F2 SL 2
## 132 N F F2 SL 3
## 133 N F F2 SL 8
## 134 N F F2 SL 10
## 135 N F F2 SL 12
## 136 N F F2 AL 1
## 137 N F F3 AL 1
## 138 N F F3 AL 9
## 139 N F F3 AL 22
## 140 N F F3 AL 3
## 141 N F F3 AL 3
## 142 N F F3 AL 5
## 143 N F F3 AL 15
## 144 N F F3 AL 18
## 145 N F F3 AL 22
## 146 N F F3 AL 37
```

이후 변수값의 의미를 살펴보았다.

```
?quine
```

추가로 결측치가 있는지 확인해본 결과, 다음과 같이 결측치는 존재하지않았다.

```
sum(is.na(absenteeism))
```

```
## [1] 0
```

Problem 2.1.

```
absenteeism$eth<-ifelse(absenteeism$eth=='A', 0, 1)
absenteeism$sex<-ifelse(absenteeism$sex=='M', 1, 0)
absenteeism$lrn<-ifelse(absenteeism$lrn=='SL', 1, 0)
absenteeism
```

```
##      X eth sex age lrn days
## 1    1  0  1  F0   1    2
## 2    2  0  1  F0   1   11
## 3    3  0  1  F0   1   14
## 4    4  0  1  F0   0    5
## 5    5  0  1  F0   0    5
## 6    6  0  1  F0   0   13
## 7    7  0  1  F0   0   20
## 8    8  0  1  F0   0   22
## 9    9  0  1  F1   1    6
## 10   10 0  1  F1   1    6
## 11   11 0  1  F1   1   15
## 12   12 0  1  F1   0    7
## 13   13 0  1  F1   0   14
## 14   14 0  1  F2   1    6
## 15   15 0  1  F2   1   32
## 16   16 0  1  F2   1   53
## 17   17 0  1  F2   1   57
## 18   18 0  1  F2   0   14
## 19   19 0  1  F2   0   16
## 20   20 0  1  F2   0   16
## 21   21 0  1  F2   0   17
## 22   22 0  1  F2   0   40
## 23   23 0  1  F2   0   43
## 24   24 0  1  F2   0   46
## 25   25 0  1  F3   0    8
## 26   26 0  1  F3   0   23
## 27   27 0  1  F3   0   23
## 28   28 0  1  F3   0   28
## 29   29 0  1  F3   0   34
## 30   30 0  1  F3   0   36
## 31   31 0  1  F3   0   38
## 32   32 0  0  F0   1    3
## 33   33 0  0  F0   0    5
## 34   34 0  0  F0   0   11
## 35   35 0  0  F0   0   24
```

## 36	36	0	0	F0	0	45
## 37	37	0	0	F1	1	5
## 38	38	0	0	F1	1	6
## 39	39	0	0	F1	1	6
## 40	40	0	0	F1	1	9
## 41	41	0	0	F1	1	13
## 42	42	0	0	F1	1	23
## 43	43	0	0	F1	1	25
## 44	44	0	0	F1	1	32
## 45	45	0	0	F1	1	53
## 46	46	0	0	F1	1	54
## 47	47	0	0	F1	0	5
## 48	48	0	0	F1	0	5
## 49	49	0	0	F1	0	11
## 50	50	0	0	F1	0	17
## 51	51	0	0	F1	0	19
## 52	52	0	0	F2	1	8
## 53	53	0	0	F2	1	13
## 54	54	0	0	F2	1	14
## 55	55	0	0	F2	1	20
## 56	56	0	0	F2	1	47
## 57	57	0	0	F2	1	48
## 58	58	0	0	F2	1	60
## 59	59	0	0	F2	1	81
## 60	60	0	0	F2	0	2
## 61	61	0	0	F3	0	0
## 62	62	0	0	F3	0	2
## 63	63	0	0	F3	0	3
## 64	64	0	0	F3	0	5
## 65	65	0	0	F3	0	10
## 66	66	0	0	F3	0	14
## 67	67	0	0	F3	0	21
## 68	68	0	0	F3	0	36
## 69	69	0	0	F3	0	40
## 70	70	1	1	F0	1	6
## 71	71	1	1	F0	1	17
## 72	72	1	1	F0	1	67
## 73	73	1	1	F0	0	0
## 74	74	1	1	F0	0	0
## 75	75	1	1	F0	0	2
## 76	76	1	1	F0	0	7

## 77	77	1	1	F0	0	11
## 78	78	1	1	F0	0	12
## 79	79	1	1	F1	1	0
## 80	80	1	1	F1	1	0
## 81	81	1	1	F1	1	5
## 82	82	1	1	F1	1	5
## 83	83	1	1	F1	1	5
## 84	84	1	1	F1	1	11
## 85	85	1	1	F1	1	17
## 86	86	1	1	F1	0	3
## 87	87	1	1	F1	0	4
## 88	88	1	1	F2	1	22
## 89	89	1	1	F2	1	30
## 90	90	1	1	F2	1	36
## 91	91	1	1	F2	0	8
## 92	92	1	1	F2	0	0
## 93	93	1	1	F2	0	1
## 94	94	1	1	F2	0	5
## 95	95	1	1	F2	0	7
## 96	96	1	1	F2	0	16
## 97	97	1	1	F2	0	27
## 98	98	1	1	F3	0	0
## 99	99	1	1	F3	0	30
## 100	100	1	1	F3	0	10
## 101	101	1	1	F3	0	14
## 102	102	1	1	F3	0	27
## 103	103	1	1	F3	0	41
## 104	104	1	1	F3	0	69
## 105	105	1	0	F0	1	25
## 106	106	1	0	F0	0	10
## 107	107	1	0	F0	0	11
## 108	108	1	0	F0	0	20
## 109	109	1	0	F0	0	33
## 110	110	1	0	F1	1	5
## 111	111	1	0	F1	1	7
## 112	112	1	0	F1	1	0
## 113	113	1	0	F1	1	1
## 114	114	1	0	F1	1	5
## 115	115	1	0	F1	1	5
## 116	116	1	0	F1	1	5
## 117	117	1	0	F1	1	5

```
## 118 118 1 0 F1 1 7
## 119 119 1 0 F1 1 11
## 120 120 1 0 F1 1 15
## 121 121 1 0 F1 0 5
## 122 122 1 0 F1 0 14
## 123 123 1 0 F1 0 6
## 124 124 1 0 F1 0 6
## 125 125 1 0 F1 0 7
## 126 126 1 0 F1 0 28
## 127 127 1 0 F2 1 0
## 128 128 1 0 F2 1 5
## 129 129 1 0 F2 1 14
## 130 130 1 0 F2 1 2
## 131 131 1 0 F2 1 2
## 132 132 1 0 F2 1 3
## 133 133 1 0 F2 1 8
## 134 134 1 0 F2 1 10
## 135 135 1 0 F2 1 12
## 136 136 1 0 F2 0 1
## 137 137 1 0 F3 0 1
## 138 138 1 0 F3 0 9
## 139 139 1 0 F3 0 22
## 140 140 1 0 F3 0 3
## 141 141 1 0 F3 0 3
## 142 142 1 0 F3 0 5
## 143 143 1 0 F3 0 15
## 144 144 1 0 F3 0 18
## 145 145 1 0 F3 0 22
## 146 146 1 0 F3 0 37
```

로 위치럼 데이터셋이 바뀌었다.

Problem 2.2.

```
attach(absenteeism)
fit2<-lm(days~eth+sex+lrn)
summary(fit2)
```

```
##
## Call:
## lm(formula = days ~ eth + sex + lrn)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.190 -10.078  -4.928   5.768  59.914
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   18.932      2.570   7.365 1.32e-11 ***
## eth           -9.112      2.599  -3.506 0.000609 ***
## sex            3.104      2.637   1.177 0.241108
## lrn            2.154      2.651   0.813 0.417732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.67 on 142 degrees of freedom
## Multiple R-squared:  0.08933,    Adjusted R-squared:  0.07009
## F-statistic: 4.643 on 3 and 142 DF,  p-value: 0.003967
```

days변수를 종속변수로 하여 문제2.1의 변수 3개로 선형회귀 적합한 결과는 위와 같다.

Problem 2.3.

```
fc2 <- fit2$coefficients
mathexpression2 <- paste("days =", paste(round(fc2[1],2), paste(round(fc2[-1],2), names(fc2[-1]), sep="
mathexpression2

## [1] "days = 18.93 + -9.11 * eth + 3.1 * sex + 2.15 * lrn + e"
```

위처럼 적합한 수학적 식이 나타난다.

적합된 값들을 해석해보면, eth, sex와 lrn이 모두 0일때는 days가 18.93으로 추정된다.

이때 sex와 lrn이 고정되어있으면, eth가 0에서 1로 변화할때 평균적으로 days가 9.11감소한다.

또한, eth와 lrn이 고정되어있으면, sex가 0에서 1로 변화할때 평균적으로 days가 3.1정도 증가한다.

마지막으로 sex와 eth이 고정되어있으면, lrn이 0에서 1로 변화할때 평균적으로 days가 2.15 증가한다.
는 의미이다.

Problem 2.4.

```
summary(fit2)["adj.r.squared"]
```

```
## $adj.r.squared
## [1] 0.07009202
```

adjusted r squared는 위와 같이 약 0.07으로 나온다.

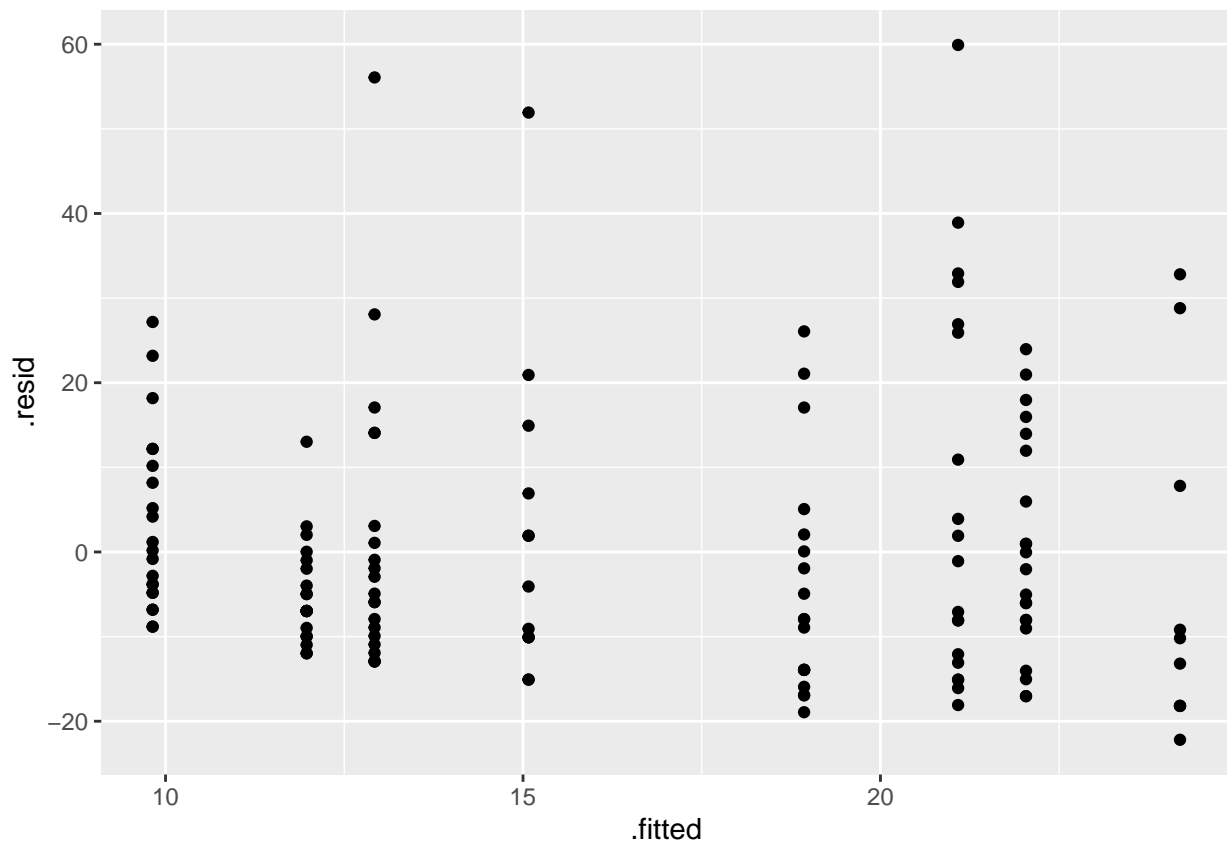
해석은 전체 변동의 평균(MST)중 모델이 설명하지 못하는 변동의 평균(MSE)의 비율이 $1-0.07=0.93$ 으로 약 93%라고 말할 수 있다. 따라서 90%이상을 설명하지 못하고 데이터를 설명할 좋은 모델이라고 볼 수 없다.

Problem 2.5.

```
inf2<-fortify(fit2, absenteeism)
names(inf2)
```

```
## [1] "X"      "eth"     "sex"     "age"     "lrn"
## [6] "days"  ".hat"    ".sigma"  ".cooksd" ".fitted"
## [11] ".resid" ".stdresid"
```

```
ggplot(inf2, aes(x = .fitted , y = .resid))+
  geom_point()
```



eth, sex와 lrn 모두가 범주형 자료이므로, fitted value는 유한개이고, $2^3=8$ 개의 적합값이 존재하게 된다.

이때 잔차들을 살펴보면, 대부분 40이하이고, 50이상은 3개이다. 또한, -20 부근의 -20이하인 점 하나를 제외하고는 모두들 -20이상이다.

따라서 잔차의 경험적 분포가 대칭적이지 않고, 전반적으로는 양수로 치워쳐있다.

그러므로, 선형회귀분석이 좋은 모델이라고 보기 어렵다.

Problem 2.6.

```
newdata1 <- tibble(eth = c(1,1,1,0,0),
sex = c(0,1,0,1,0),
lrn = c(0,0,1,1,0))
newdata1
```

```
## # A tibble: 5 x 3
##   eth   sex   lrn
##   <dbl> <dbl> <dbl>
## 1     1     0     0
## 2     1     1     0
## 3     1     0     1
## 4     0     1     1
## 5     0     0     0
```

이제 5번에서 사용한 선형회귀모델을 이용하여 나온 예측값을 구해보면 다음과 같다.

```
result2<-predict(fit2,newdata = newdata1)
result2
```

```
##           1           2           3           4           5
## 9.819607 12.923862 11.973764 24.190261 18.931848
```

위의 테이블에 함께 나타내면 다음과 같다.

```
newdata1%>%mutate(result2)
```

```
## # A tibble: 5 x 4
##   eth   sex   lrn result2
##   <dbl> <dbl> <dbl>   <dbl>
## 1     1     0     0     9.82
## 2     1     1     0    12.9
## 3     1     0     1    12.0
## 4     0     1     1    24.2
## 5     0     0     0    18.9
```

```
detach(absenteeism)
```

0.1.3 Problem 3.

true regression ftn이 아래와 같이 $f(x)=x^2$ 이다.

또한, 아래와 같이 x,y 를 랜덤하게 100개의 자료로 만들어서 살펴본다.

```
f = function(x) { x^2}
get_sim_data = function(f, sample_size = 100) {
x = runif(n = sample_size, min = 0, max = 1)
```

```

y = rnorm(n = sample_size, mean = f(x), sd = 0.3)
data.frame(x, y)
}

```

먼저, seed=1로 고정시켜놓고, 4가지 모델에 적합시킨다.

```

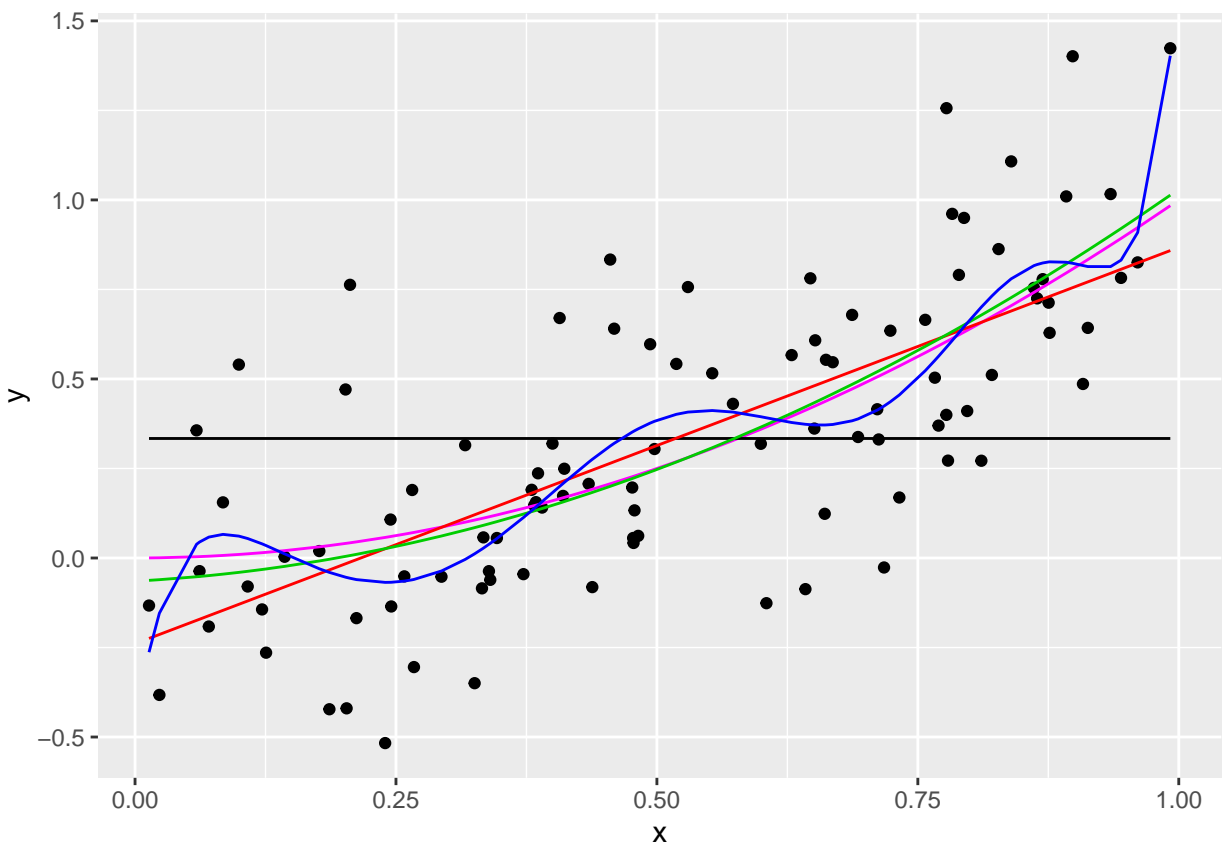
set.seed(1)
sim_data = get_sim_data(f)
fit_0 = lm(y ~ 1, data = sim_data)
fit_1 = lm(y ~ poly(x, degree = 1), data = sim_data)
fit_2 = lm(y ~ poly(x, degree = 2), data = sim_data)
fit_9 = lm(y ~ poly(x, degree = 9), data = sim_data)

```

```

sim_data1<-sim_data%>%mutate(fit0=fit_0$fitted.values,fit1=fit_1$fitted.values,fit2=fit_2$fitted.values)
ggplot(sim_data1,mapping = aes(x = x, y = y)) +
  geom_point(mapping = aes(x = x, y = y))+
  geom_line(mapping = aes(x = x, y = TRUE1),color=6)+
  geom_line(mapping = aes(x = x, y = fit0),color=1)+
  geom_line(mapping = aes(x = x, y = fit1),color=2)+
  geom_line(mapping = aes(x = x, y = fit2),color=3)+
  geom_line(mapping = aes(x = x, y = fit9),color=4)

```



```
labs(
  title = "the true and four polynomial fitted lines"
)
```

```
## $title
## [1] "the true and four polynomial fitted lines"
##
## attr(,"class")
## [1] "labels"
```

위처럼 데이터 scatter plot 위에 실제 회귀선과 각각 적합시킨 선을 그렸다.

degree가 0인 것은 데이터를 잘 설명하지 못하고 전체 평균만 보여주고 있다.

degree가 9인 것을 제외하고, 다른 0아닌 degree를 갖는 모형들은 서로 비슷한 회귀선을 보여준다.

degree가 9인 것은 추세가 있는 주기함수처럼 나왔다.

이제 250번 시뮬레이션을 돌려서, box plot을 그려보자.

```
?geom_boxplot
```

```
set.seed(1)
n_sims = 250
n_models = 4
x = data.frame(x = 0.90)
predictions = data.frame( degree=c(rep("0",250),rep("1",250),rep("2",250),rep("9",250)), pred=c(1:1000)
predictions
```

```
##      degree pred
## 1         0    1
## 2         0    2
## 3         0    3
## 4         0    4
## 5         0    5
## 6         0    6
## 7         0    7
## 8         0    8
## 9         0    9
## 10        0   10
## 11        0   11
## 12        0   12
## 13        0   13
## 14        0   14
## 15        0   15
```

## 16	0	16
## 17	0	17
## 18	0	18
## 19	0	19
## 20	0	20
## 21	0	21
## 22	0	22
## 23	0	23
## 24	0	24
## 25	0	25
## 26	0	26
## 27	0	27
## 28	0	28
## 29	0	29
## 30	0	30
## 31	0	31
## 32	0	32
## 33	0	33
## 34	0	34
## 35	0	35
## 36	0	36
## 37	0	37
## 38	0	38
## 39	0	39
## 40	0	40
## 41	0	41
## 42	0	42
## 43	0	43
## 44	0	44
## 45	0	45
## 46	0	46
## 47	0	47
## 48	0	48
## 49	0	49
## 50	0	50
## 51	0	51
## 52	0	52
## 53	0	53
## 54	0	54
## 55	0	55
## 56	0	56

## 57	0	57
## 58	0	58
## 59	0	59
## 60	0	60
## 61	0	61
## 62	0	62
## 63	0	63
## 64	0	64
## 65	0	65
## 66	0	66
## 67	0	67
## 68	0	68
## 69	0	69
## 70	0	70
## 71	0	71
## 72	0	72
## 73	0	73
## 74	0	74
## 75	0	75
## 76	0	76
## 77	0	77
## 78	0	78
## 79	0	79
## 80	0	80
## 81	0	81
## 82	0	82
## 83	0	83
## 84	0	84
## 85	0	85
## 86	0	86
## 87	0	87
## 88	0	88
## 89	0	89
## 90	0	90
## 91	0	91
## 92	0	92
## 93	0	93
## 94	0	94
## 95	0	95
## 96	0	96
## 97	0	97

## 98	0	98
## 99	0	99
## 100	0	100
## 101	0	101
## 102	0	102
## 103	0	103
## 104	0	104
## 105	0	105
## 106	0	106
## 107	0	107
## 108	0	108
## 109	0	109
## 110	0	110
## 111	0	111
## 112	0	112
## 113	0	113
## 114	0	114
## 115	0	115
## 116	0	116
## 117	0	117
## 118	0	118
## 119	0	119
## 120	0	120
## 121	0	121
## 122	0	122
## 123	0	123
## 124	0	124
## 125	0	125
## 126	0	126
## 127	0	127
## 128	0	128
## 129	0	129
## 130	0	130
## 131	0	131
## 132	0	132
## 133	0	133
## 134	0	134
## 135	0	135
## 136	0	136
## 137	0	137
## 138	0	138

## 139	0	139
## 140	0	140
## 141	0	141
## 142	0	142
## 143	0	143
## 144	0	144
## 145	0	145
## 146	0	146
## 147	0	147
## 148	0	148
## 149	0	149
## 150	0	150
## 151	0	151
## 152	0	152
## 153	0	153
## 154	0	154
## 155	0	155
## 156	0	156
## 157	0	157
## 158	0	158
## 159	0	159
## 160	0	160
## 161	0	161
## 162	0	162
## 163	0	163
## 164	0	164
## 165	0	165
## 166	0	166
## 167	0	167
## 168	0	168
## 169	0	169
## 170	0	170
## 171	0	171
## 172	0	172
## 173	0	173
## 174	0	174
## 175	0	175
## 176	0	176
## 177	0	177
## 178	0	178
## 179	0	179

## 180	0	180
## 181	0	181
## 182	0	182
## 183	0	183
## 184	0	184
## 185	0	185
## 186	0	186
## 187	0	187
## 188	0	188
## 189	0	189
## 190	0	190
## 191	0	191
## 192	0	192
## 193	0	193
## 194	0	194
## 195	0	195
## 196	0	196
## 197	0	197
## 198	0	198
## 199	0	199
## 200	0	200
## 201	0	201
## 202	0	202
## 203	0	203
## 204	0	204
## 205	0	205
## 206	0	206
## 207	0	207
## 208	0	208
## 209	0	209
## 210	0	210
## 211	0	211
## 212	0	212
## 213	0	213
## 214	0	214
## 215	0	215
## 216	0	216
## 217	0	217
## 218	0	218
## 219	0	219
## 220	0	220

## 221	0	221
## 222	0	222
## 223	0	223
## 224	0	224
## 225	0	225
## 226	0	226
## 227	0	227
## 228	0	228
## 229	0	229
## 230	0	230
## 231	0	231
## 232	0	232
## 233	0	233
## 234	0	234
## 235	0	235
## 236	0	236
## 237	0	237
## 238	0	238
## 239	0	239
## 240	0	240
## 241	0	241
## 242	0	242
## 243	0	243
## 244	0	244
## 245	0	245
## 246	0	246
## 247	0	247
## 248	0	248
## 249	0	249
## 250	0	250
## 251	1	251
## 252	1	252
## 253	1	253
## 254	1	254
## 255	1	255
## 256	1	256
## 257	1	257
## 258	1	258
## 259	1	259
## 260	1	260
## 261	1	261

## 262	1	262
## 263	1	263
## 264	1	264
## 265	1	265
## 266	1	266
## 267	1	267
## 268	1	268
## 269	1	269
## 270	1	270
## 271	1	271
## 272	1	272
## 273	1	273
## 274	1	274
## 275	1	275
## 276	1	276
## 277	1	277
## 278	1	278
## 279	1	279
## 280	1	280
## 281	1	281
## 282	1	282
## 283	1	283
## 284	1	284
## 285	1	285
## 286	1	286
## 287	1	287
## 288	1	288
## 289	1	289
## 290	1	290
## 291	1	291
## 292	1	292
## 293	1	293
## 294	1	294
## 295	1	295
## 296	1	296
## 297	1	297
## 298	1	298
## 299	1	299
## 300	1	300
## 301	1	301
## 302	1	302

## 303	1	303
## 304	1	304
## 305	1	305
## 306	1	306
## 307	1	307
## 308	1	308
## 309	1	309
## 310	1	310
## 311	1	311
## 312	1	312
## 313	1	313
## 314	1	314
## 315	1	315
## 316	1	316
## 317	1	317
## 318	1	318
## 319	1	319
## 320	1	320
## 321	1	321
## 322	1	322
## 323	1	323
## 324	1	324
## 325	1	325
## 326	1	326
## 327	1	327
## 328	1	328
## 329	1	329
## 330	1	330
## 331	1	331
## 332	1	332
## 333	1	333
## 334	1	334
## 335	1	335
## 336	1	336
## 337	1	337
## 338	1	338
## 339	1	339
## 340	1	340
## 341	1	341
## 342	1	342
## 343	1	343

## 344	1	344
## 345	1	345
## 346	1	346
## 347	1	347
## 348	1	348
## 349	1	349
## 350	1	350
## 351	1	351
## 352	1	352
## 353	1	353
## 354	1	354
## 355	1	355
## 356	1	356
## 357	1	357
## 358	1	358
## 359	1	359
## 360	1	360
## 361	1	361
## 362	1	362
## 363	1	363
## 364	1	364
## 365	1	365
## 366	1	366
## 367	1	367
## 368	1	368
## 369	1	369
## 370	1	370
## 371	1	371
## 372	1	372
## 373	1	373
## 374	1	374
## 375	1	375
## 376	1	376
## 377	1	377
## 378	1	378
## 379	1	379
## 380	1	380
## 381	1	381
## 382	1	382
## 383	1	383
## 384	1	384

## 385	1	385
## 386	1	386
## 387	1	387
## 388	1	388
## 389	1	389
## 390	1	390
## 391	1	391
## 392	1	392
## 393	1	393
## 394	1	394
## 395	1	395
## 396	1	396
## 397	1	397
## 398	1	398
## 399	1	399
## 400	1	400
## 401	1	401
## 402	1	402
## 403	1	403
## 404	1	404
## 405	1	405
## 406	1	406
## 407	1	407
## 408	1	408
## 409	1	409
## 410	1	410
## 411	1	411
## 412	1	412
## 413	1	413
## 414	1	414
## 415	1	415
## 416	1	416
## 417	1	417
## 418	1	418
## 419	1	419
## 420	1	420
## 421	1	421
## 422	1	422
## 423	1	423
## 424	1	424
## 425	1	425

## 426	1	426
## 427	1	427
## 428	1	428
## 429	1	429
## 430	1	430
## 431	1	431
## 432	1	432
## 433	1	433
## 434	1	434
## 435	1	435
## 436	1	436
## 437	1	437
## 438	1	438
## 439	1	439
## 440	1	440
## 441	1	441
## 442	1	442
## 443	1	443
## 444	1	444
## 445	1	445
## 446	1	446
## 447	1	447
## 448	1	448
## 449	1	449
## 450	1	450
## 451	1	451
## 452	1	452
## 453	1	453
## 454	1	454
## 455	1	455
## 456	1	456
## 457	1	457
## 458	1	458
## 459	1	459
## 460	1	460
## 461	1	461
## 462	1	462
## 463	1	463
## 464	1	464
## 465	1	465
## 466	1	466

## 467	1	467
## 468	1	468
## 469	1	469
## 470	1	470
## 471	1	471
## 472	1	472
## 473	1	473
## 474	1	474
## 475	1	475
## 476	1	476
## 477	1	477
## 478	1	478
## 479	1	479
## 480	1	480
## 481	1	481
## 482	1	482
## 483	1	483
## 484	1	484
## 485	1	485
## 486	1	486
## 487	1	487
## 488	1	488
## 489	1	489
## 490	1	490
## 491	1	491
## 492	1	492
## 493	1	493
## 494	1	494
## 495	1	495
## 496	1	496
## 497	1	497
## 498	1	498
## 499	1	499
## 500	1	500
## 501	2	501
## 502	2	502
## 503	2	503
## 504	2	504
## 505	2	505
## 506	2	506
## 507	2	507

## 508	2	508
## 509	2	509
## 510	2	510
## 511	2	511
## 512	2	512
## 513	2	513
## 514	2	514
## 515	2	515
## 516	2	516
## 517	2	517
## 518	2	518
## 519	2	519
## 520	2	520
## 521	2	521
## 522	2	522
## 523	2	523
## 524	2	524
## 525	2	525
## 526	2	526
## 527	2	527
## 528	2	528
## 529	2	529
## 530	2	530
## 531	2	531
## 532	2	532
## 533	2	533
## 534	2	534
## 535	2	535
## 536	2	536
## 537	2	537
## 538	2	538
## 539	2	539
## 540	2	540
## 541	2	541
## 542	2	542
## 543	2	543
## 544	2	544
## 545	2	545
## 546	2	546
## 547	2	547
## 548	2	548

## 549	2	549
## 550	2	550
## 551	2	551
## 552	2	552
## 553	2	553
## 554	2	554
## 555	2	555
## 556	2	556
## 557	2	557
## 558	2	558
## 559	2	559
## 560	2	560
## 561	2	561
## 562	2	562
## 563	2	563
## 564	2	564
## 565	2	565
## 566	2	566
## 567	2	567
## 568	2	568
## 569	2	569
## 570	2	570
## 571	2	571
## 572	2	572
## 573	2	573
## 574	2	574
## 575	2	575
## 576	2	576
## 577	2	577
## 578	2	578
## 579	2	579
## 580	2	580
## 581	2	581
## 582	2	582
## 583	2	583
## 584	2	584
## 585	2	585
## 586	2	586
## 587	2	587
## 588	2	588
## 589	2	589

## 590	2	590
## 591	2	591
## 592	2	592
## 593	2	593
## 594	2	594
## 595	2	595
## 596	2	596
## 597	2	597
## 598	2	598
## 599	2	599
## 600	2	600
## 601	2	601
## 602	2	602
## 603	2	603
## 604	2	604
## 605	2	605
## 606	2	606
## 607	2	607
## 608	2	608
## 609	2	609
## 610	2	610
## 611	2	611
## 612	2	612
## 613	2	613
## 614	2	614
## 615	2	615
## 616	2	616
## 617	2	617
## 618	2	618
## 619	2	619
## 620	2	620
## 621	2	621
## 622	2	622
## 623	2	623
## 624	2	624
## 625	2	625
## 626	2	626
## 627	2	627
## 628	2	628
## 629	2	629
## 630	2	630

## 631	2	631
## 632	2	632
## 633	2	633
## 634	2	634
## 635	2	635
## 636	2	636
## 637	2	637
## 638	2	638
## 639	2	639
## 640	2	640
## 641	2	641
## 642	2	642
## 643	2	643
## 644	2	644
## 645	2	645
## 646	2	646
## 647	2	647
## 648	2	648
## 649	2	649
## 650	2	650
## 651	2	651
## 652	2	652
## 653	2	653
## 654	2	654
## 655	2	655
## 656	2	656
## 657	2	657
## 658	2	658
## 659	2	659
## 660	2	660
## 661	2	661
## 662	2	662
## 663	2	663
## 664	2	664
## 665	2	665
## 666	2	666
## 667	2	667
## 668	2	668
## 669	2	669
## 670	2	670
## 671	2	671

## 672	2	672
## 673	2	673
## 674	2	674
## 675	2	675
## 676	2	676
## 677	2	677
## 678	2	678
## 679	2	679
## 680	2	680
## 681	2	681
## 682	2	682
## 683	2	683
## 684	2	684
## 685	2	685
## 686	2	686
## 687	2	687
## 688	2	688
## 689	2	689
## 690	2	690
## 691	2	691
## 692	2	692
## 693	2	693
## 694	2	694
## 695	2	695
## 696	2	696
## 697	2	697
## 698	2	698
## 699	2	699
## 700	2	700
## 701	2	701
## 702	2	702
## 703	2	703
## 704	2	704
## 705	2	705
## 706	2	706
## 707	2	707
## 708	2	708
## 709	2	709
## 710	2	710
## 711	2	711
## 712	2	712

## 713	2	713
## 714	2	714
## 715	2	715
## 716	2	716
## 717	2	717
## 718	2	718
## 719	2	719
## 720	2	720
## 721	2	721
## 722	2	722
## 723	2	723
## 724	2	724
## 725	2	725
## 726	2	726
## 727	2	727
## 728	2	728
## 729	2	729
## 730	2	730
## 731	2	731
## 732	2	732
## 733	2	733
## 734	2	734
## 735	2	735
## 736	2	736
## 737	2	737
## 738	2	738
## 739	2	739
## 740	2	740
## 741	2	741
## 742	2	742
## 743	2	743
## 744	2	744
## 745	2	745
## 746	2	746
## 747	2	747
## 748	2	748
## 749	2	749
## 750	2	750
## 751	9	751
## 752	9	752
## 753	9	753

## 754	9	754
## 755	9	755
## 756	9	756
## 757	9	757
## 758	9	758
## 759	9	759
## 760	9	760
## 761	9	761
## 762	9	762
## 763	9	763
## 764	9	764
## 765	9	765
## 766	9	766
## 767	9	767
## 768	9	768
## 769	9	769
## 770	9	770
## 771	9	771
## 772	9	772
## 773	9	773
## 774	9	774
## 775	9	775
## 776	9	776
## 777	9	777
## 778	9	778
## 779	9	779
## 780	9	780
## 781	9	781
## 782	9	782
## 783	9	783
## 784	9	784
## 785	9	785
## 786	9	786
## 787	9	787
## 788	9	788
## 789	9	789
## 790	9	790
## 791	9	791
## 792	9	792
## 793	9	793
## 794	9	794

## 795	9	795
## 796	9	796
## 797	9	797
## 798	9	798
## 799	9	799
## 800	9	800
## 801	9	801
## 802	9	802
## 803	9	803
## 804	9	804
## 805	9	805
## 806	9	806
## 807	9	807
## 808	9	808
## 809	9	809
## 810	9	810
## 811	9	811
## 812	9	812
## 813	9	813
## 814	9	814
## 815	9	815
## 816	9	816
## 817	9	817
## 818	9	818
## 819	9	819
## 820	9	820
## 821	9	821
## 822	9	822
## 823	9	823
## 824	9	824
## 825	9	825
## 826	9	826
## 827	9	827
## 828	9	828
## 829	9	829
## 830	9	830
## 831	9	831
## 832	9	832
## 833	9	833
## 834	9	834
## 835	9	835

## 836	9	836
## 837	9	837
## 838	9	838
## 839	9	839
## 840	9	840
## 841	9	841
## 842	9	842
## 843	9	843
## 844	9	844
## 845	9	845
## 846	9	846
## 847	9	847
## 848	9	848
## 849	9	849
## 850	9	850
## 851	9	851
## 852	9	852
## 853	9	853
## 854	9	854
## 855	9	855
## 856	9	856
## 857	9	857
## 858	9	858
## 859	9	859
## 860	9	860
## 861	9	861
## 862	9	862
## 863	9	863
## 864	9	864
## 865	9	865
## 866	9	866
## 867	9	867
## 868	9	868
## 869	9	869
## 870	9	870
## 871	9	871
## 872	9	872
## 873	9	873
## 874	9	874
## 875	9	875
## 876	9	876

## 877	9	877
## 878	9	878
## 879	9	879
## 880	9	880
## 881	9	881
## 882	9	882
## 883	9	883
## 884	9	884
## 885	9	885
## 886	9	886
## 887	9	887
## 888	9	888
## 889	9	889
## 890	9	890
## 891	9	891
## 892	9	892
## 893	9	893
## 894	9	894
## 895	9	895
## 896	9	896
## 897	9	897
## 898	9	898
## 899	9	899
## 900	9	900
## 901	9	901
## 902	9	902
## 903	9	903
## 904	9	904
## 905	9	905
## 906	9	906
## 907	9	907
## 908	9	908
## 909	9	909
## 910	9	910
## 911	9	911
## 912	9	912
## 913	9	913
## 914	9	914
## 915	9	915
## 916	9	916
## 917	9	917

## 918	9	918
## 919	9	919
## 920	9	920
## 921	9	921
## 922	9	922
## 923	9	923
## 924	9	924
## 925	9	925
## 926	9	926
## 927	9	927
## 928	9	928
## 929	9	929
## 930	9	930
## 931	9	931
## 932	9	932
## 933	9	933
## 934	9	934
## 935	9	935
## 936	9	936
## 937	9	937
## 938	9	938
## 939	9	939
## 940	9	940
## 941	9	941
## 942	9	942
## 943	9	943
## 944	9	944
## 945	9	945
## 946	9	946
## 947	9	947
## 948	9	948
## 949	9	949
## 950	9	950
## 951	9	951
## 952	9	952
## 953	9	953
## 954	9	954
## 955	9	955
## 956	9	956
## 957	9	957
## 958	9	958

## 959	9	959
## 960	9	960
## 961	9	961
## 962	9	962
## 963	9	963
## 964	9	964
## 965	9	965
## 966	9	966
## 967	9	967
## 968	9	968
## 969	9	969
## 970	9	970
## 971	9	971
## 972	9	972
## 973	9	973
## 974	9	974
## 975	9	975
## 976	9	976
## 977	9	977
## 978	9	978
## 979	9	979
## 980	9	980
## 981	9	981
## 982	9	982
## 983	9	983
## 984	9	984
## 985	9	985
## 986	9	986
## 987	9	987
## 988	9	988
## 989	9	989
## 990	9	990
## 991	9	991
## 992	9	992
## 993	9	993
## 994	9	994
## 995	9	995
## 996	9	996
## 997	9	997
## 998	9	998
## 999	9	999

```
## 1000      9 1000
```

```
for(sim in 1:n_sims) {  
  sim_data = get_sim_data(f)  
  fit_0 = lm(y ~ 1, data = sim_data)  
  fit_1 = lm(y ~ poly(x, degree = 1), data = sim_data)  
  fit_2 = lm(y ~ poly(x, degree = 2), data = sim_data)  
  fit_9 = lm(y ~ poly(x, degree = 9), data = sim_data)  
  predictions[sim, 2] <- predict(fit_0, x)  
  predictions[250+sim, 2] = predict(fit_1, x)  
  predictions[500+sim, 2] = predict(fit_2, x)  
  predictions[750+sim, 2] = predict(fit_9, x)  
}  
predictions
```

##	degree	pred
## 1	0	0.3337773
## 2	0	0.3450969
## 3	0	0.3275432
## 4	0	0.3135135
## 5	0	0.2951493
## 6	0	0.3789756
## 7	0	0.3275696
## 8	0	0.2824191
## 9	0	0.3881284
## 10	0	0.3137406
## 11	0	0.3272482
## 12	0	0.2911202
## 13	0	0.3631599
## 14	0	0.2716200
## 15	0	0.3409671
## 16	0	0.3791002
## 17	0	0.3261282
## 18	0	0.3604506
## 19	0	0.3915693
## 20	0	0.3874948
## 21	0	0.4244122
## 22	0	0.3899748
## 23	0	0.3772772
## 24	0	0.3510727
## 25	0	0.2619259
## 26	0	0.3070249

## 27	0 0.2834751
## 28	0 0.2870605
## 29	0 0.3490363
## 30	0 0.3590533
## 31	0 0.3239120
## 32	0 0.3424262
## 33	0 0.3095317
## 34	0 0.2669277
## 35	0 0.3514918
## 36	0 0.3674154
## 37	0 0.3399012
## 38	0 0.3205194
## 39	0 0.3560876
## 40	0 0.3375859
## 41	0 0.3350582
## 42	0 0.2830571
## 43	0 0.3722079
## 44	0 0.2907369
## 45	0 0.4187829
## 46	0 0.2749396
## 47	0 0.2650805
## 48	0 0.3499900
## 49	0 0.3058487
## 50	0 0.3613947
## 51	0 0.3369525
## 52	0 0.3426023
## 53	0 0.3493576
## 54	0 0.2940621
## 55	0 0.3506226
## 56	0 0.3052723
## 57	0 0.2110160
## 58	0 0.3108850
## 59	0 0.3642174
## 60	0 0.3955829
## 61	0 0.3199845
## 62	0 0.4198511
## 63	0 0.3401038
## 64	0 0.3364559
## 65	0 0.3471835
## 66	0 0.2913152
## 67	0 0.3786456

## 68	0 0.3452679
## 69	0 0.3261911
## 70	0 0.3516364
## 71	0 0.3125916
## 72	0 0.3251224
## 73	0 0.3705132
## 74	0 0.2897031
## 75	0 0.3290923
## 76	0 0.2959993
## 77	0 0.2993202
## 78	0 0.3597357
## 79	0 0.3247098
## 80	0 0.3229759
## 81	0 0.3260538
## 82	0 0.2516076
## 83	0 0.3322793
## 84	0 0.2923259
## 85	0 0.3689234
## 86	0 0.2665366
## 87	0 0.2796165
## 88	0 0.3792210
## 89	0 0.3102287
## 90	0 0.3151192
## 91	0 0.3891400
## 92	0 0.3379672
## 93	0 0.3490904
## 94	0 0.3657071
## 95	0 0.3172478
## 96	0 0.3374627
## 97	0 0.2345313
## 98	0 0.3425412
## 99	0 0.3015253
## 100	0 0.2988878
## 101	0 0.3249876
## 102	0 0.3303852
## 103	0 0.3257142
## 104	0 0.3642582
## 105	0 0.3421556
## 106	0 0.3760424
## 107	0 0.2932622
## 108	0 0.3668219

## 109	0 0.3238082
## 110	0 0.3211425
## 111	0 0.2980472
## 112	0 0.3807031
## 113	0 0.4135592
## 114	0 0.3269793
## 115	0 0.3788407
## 116	0 0.3711929
## 117	0 0.3375975
## 118	0 0.3870794
## 119	0 0.3276461
## 120	0 0.3020224
## 121	0 0.3219769
## 122	0 0.3156659
## 123	0 0.2979737
## 124	0 0.3050181
## 125	0 0.3261844
## 126	0 0.3373174
## 127	0 0.3220104
## 128	0 0.2677206
## 129	0 0.3018024
## 130	0 0.3946809
## 131	0 0.3883230
## 132	0 0.3223680
## 133	0 0.3334574
## 134	0 0.3204794
## 135	0 0.3876156
## 136	0 0.4002544
## 137	0 0.3646499
## 138	0 0.3599705
## 139	0 0.3111540
## 140	0 0.4249578
## 141	0 0.3778340
## 142	0 0.3362511
## 143	0 0.3836836
## 144	0 0.3072608
## 145	0 0.3778749
## 146	0 0.2804676
## 147	0 0.3006830
## 148	0 0.2860804
## 149	0 0.3686078

## 150	0 0.3208363
## 151	0 0.3008042
## 152	0 0.3373318
## 153	0 0.3402545
## 154	0 0.3549298
## 155	0 0.3364574
## 156	0 0.2743733
## 157	0 0.3662043
## 158	0 0.3366591
## 159	0 0.3509384
## 160	0 0.4364600
## 161	0 0.3955739
## 162	0 0.3282844
## 163	0 0.2705652
## 164	0 0.3584886
## 165	0 0.3446348
## 166	0 0.3560046
## 167	0 0.3768637
## 168	0 0.3555927
## 169	0 0.3284809
## 170	0 0.3570923
## 171	0 0.3259607
## 172	0 0.3515947
## 173	0 0.2583245
## 174	0 0.2730724
## 175	0 0.3143965
## 176	0 0.4211369
## 177	0 0.3079266
## 178	0 0.3138636
## 179	0 0.3469282
## 180	0 0.3560347
## 181	0 0.2915779
## 182	0 0.3753665
## 183	0 0.3285501
## 184	0 0.4341287
## 185	0 0.3036564
## 186	0 0.3440323
## 187	0 0.3324324
## 188	0 0.3564183
## 189	0 0.3233259
## 190	0 0.4121806

## 191	0 0.3368967
## 192	0 0.3657588
## 193	0 0.2381400
## 194	0 0.3110714
## 195	0 0.2925491
## 196	0 0.4195651
## 197	0 0.2970093
## 198	0 0.2769279
## 199	0 0.3254057
## 200	0 0.3283259
## 201	0 0.3990308
## 202	0 0.3560862
## 203	0 0.3752261
## 204	0 0.3005194
## 205	0 0.3411774
## 206	0 0.3787099
## 207	0 0.3574203
## 208	0 0.3157656
## 209	0 0.3496920
## 210	0 0.3286774
## 211	0 0.3750230
## 212	0 0.4256602
## 213	0 0.3845162
## 214	0 0.3318941
## 215	0 0.3733874
## 216	0 0.3343319
## 217	0 0.3208344
## 218	0 0.3190963
## 219	0 0.2918989
## 220	0 0.3053383
## 221	0 0.3243654
## 222	0 0.3587505
## 223	0 0.3317574
## 224	0 0.3164904
## 225	0 0.3056071
## 226	0 0.3189625
## 227	0 0.3211595
## 228	0 0.3006038
## 229	0 0.3280832
## 230	0 0.3786841
## 231	0 0.3061049

## 232	0 0.3765062
## 233	0 0.3452192
## 234	0 0.2550093
## 235	0 0.4042634
## 236	0 0.3302657
## 237	0 0.2675331
## 238	0 0.3269612
## 239	0 0.3997008
## 240	0 0.3531942
## 241	0 0.3379427
## 242	0 0.3647217
## 243	0 0.3988687
## 244	0 0.3954203
## 245	0 0.2453339
## 246	0 0.2326380
## 247	0 0.3916927
## 248	0 0.3259835
## 249	0 0.3164998
## 250	0 0.3998042
## 251	1 0.7568053
## 252	1 0.7791787
## 253	1 0.7986245
## 254	1 0.7057814
## 255	1 0.7014001
## 256	1 0.7124436
## 257	1 0.7171964
## 258	1 0.7187661
## 259	1 0.8373416
## 260	1 0.7482137
## 261	1 0.7006589
## 262	1 0.6984280
## 263	1 0.7346525
## 264	1 0.6625485
## 265	1 0.7776111
## 266	1 0.7728981
## 267	1 0.7340069
## 268	1 0.7284181
## 269	1 0.8904905
## 270	1 0.7552716
## 271	1 0.8808816
## 272	1 0.8078457

## 273	1 0.7562547
## 274	1 0.8425114
## 275	1 0.8047761
## 276	1 0.7865805
## 277	1 0.8105835
## 278	1 0.6572877
## 279	1 0.8694137
## 280	1 0.7492678
## 281	1 0.7867351
## 282	1 0.7778537
## 283	1 0.7035251
## 284	1 0.6948606
## 285	1 0.7168197
## 286	1 0.7650125
## 287	1 0.7059300
## 288	1 0.6905378
## 289	1 0.7117988
## 290	1 0.7548611
## 291	1 0.7529574
## 292	1 0.7647196
## 293	1 0.7395792
## 294	1 0.6350725
## 295	1 0.7954324
## 296	1 0.6355274
## 297	1 0.6311664
## 298	1 0.7661664
## 299	1 0.7572481
## 300	1 0.7929681
## 301	1 0.6833916
## 302	1 0.7089589
## 303	1 0.8073307
## 304	1 0.7049328
## 305	1 0.8172775
## 306	1 0.6692351
## 307	1 0.6518120
## 308	1 0.6980545
## 309	1 0.7560349
## 310	1 0.7397632
## 311	1 0.7270582
## 312	1 0.7870869
## 313	1 0.7458085

## 314	1 0.7399268
## 315	1 0.7986991
## 316	1 0.6811552
## 317	1 0.7169152
## 318	1 0.7921494
## 319	1 0.7712099
## 320	1 0.6399682
## 321	1 0.7524928
## 322	1 0.7153157
## 323	1 0.7941604
## 324	1 0.6691938
## 325	1 0.7636558
## 326	1 0.6622045
## 327	1 0.7279994
## 328	1 0.7497233
## 329	1 0.6492201
## 330	1 0.6592321
## 331	1 0.7375808
## 332	1 0.7291106
## 333	1 0.7168532
## 334	1 0.7744579
## 335	1 0.6539675
## 336	1 0.6979727
## 337	1 0.7340101
## 338	1 0.7295753
## 339	1 0.7117964
## 340	1 0.6564926
## 341	1 0.7322955
## 342	1 0.7276242
## 343	1 0.7699248
## 344	1 0.7074849
## 345	1 0.7533456
## 346	1 0.7372030
## 347	1 0.5922660
## 348	1 0.7434254
## 349	1 0.6811489
## 350	1 0.7922025
## 351	1 0.7843930
## 352	1 0.7642449
## 353	1 0.7978406
## 354	1 0.7685771

## 355	1 0.6785283
## 356	1 0.7078658
## 357	1 0.6093342
## 358	1 0.8140085
## 359	1 0.7127664
## 360	1 0.7167010
## 361	1 0.6896506
## 362	1 0.7472204
## 363	1 0.8119390
## 364	1 0.6819788
## 365	1 0.7667466
## 366	1 0.7170539
## 367	1 0.7629402
## 368	1 0.7716024
## 369	1 0.7535801
## 370	1 0.6523713
## 371	1 0.6885349
## 372	1 0.7187246
## 373	1 0.7445296
## 374	1 0.7064256
## 375	1 0.6860830
## 376	1 0.6964242
## 377	1 0.6822130
## 378	1 0.6810267
## 379	1 0.6939829
## 380	1 0.8154077
## 381	1 0.7235181
## 382	1 0.7434917
## 383	1 0.7758201
## 384	1 0.7006511
## 385	1 0.7990525
## 386	1 0.8698472
## 387	1 0.7974605
## 388	1 0.6384634
## 389	1 0.7022351
## 390	1 0.8320430
## 391	1 0.7775866
## 392	1 0.6766380
## 393	1 0.7164725
## 394	1 0.7583983
## 395	1 0.7659006

## 396	1 0.7809538
## 397	1 0.7571387
## 398	1 0.7146607
## 399	1 0.8133842
## 400	1 0.7050464
## 401	1 0.6822188
## 402	1 0.8168247
## 403	1 0.6541412
## 404	1 0.7353419
## 405	1 0.7489258
## 406	1 0.7329911
## 407	1 0.8294194
## 408	1 0.7002462
## 409	1 0.7117836
## 410	1 0.8533582
## 411	1 0.7986531
## 412	1 0.6849350
## 413	1 0.7289937
## 414	1 0.6963829
## 415	1 0.7205154
## 416	1 0.7321048
## 417	1 0.7967292
## 418	1 0.6643262
## 419	1 0.7270613
## 420	1 0.7062113
## 421	1 0.7555178
## 422	1 0.8192360
## 423	1 0.6345486
## 424	1 0.6382622
## 425	1 0.7956840
## 426	1 0.8571828
## 427	1 0.7369852
## 428	1 0.6547886
## 429	1 0.8453996
## 430	1 0.7253823
## 431	1 0.7355830
## 432	1 0.8125894
## 433	1 0.7499391
## 434	1 0.8196270
## 435	1 0.7782712
## 436	1 0.7301320

## 437	1 0.7692665
## 438	1 0.7787152
## 439	1 0.7299725
## 440	1 0.7848731
## 441	1 0.7426395
## 442	1 0.6861826
## 443	1 0.6618762
## 444	1 0.7242009
## 445	1 0.7108807
## 446	1 0.7943623
## 447	1 0.7951971
## 448	1 0.7079652
## 449	1 0.7445977
## 450	1 0.7455928
## 451	1 0.7831511
## 452	1 0.8507183
## 453	1 0.7638984
## 454	1 0.7033144
## 455	1 0.7257080
## 456	1 0.7594407
## 457	1 0.6633640
## 458	1 0.7650146
## 459	1 0.7442306
## 460	1 0.7046335
## 461	1 0.7872159
## 462	1 0.7970025
## 463	1 0.8033865
## 464	1 0.6674764
## 465	1 0.7952056
## 466	1 0.7841541
## 467	1 0.7756470
## 468	1 0.7767714
## 469	1 0.6820115
## 470	1 0.7492906
## 471	1 0.7776523
## 472	1 0.7379818
## 473	1 0.6255741
## 474	1 0.7657130
## 475	1 0.6339855
## 476	1 0.6369303
## 477	1 0.6935852

## 478	1 0.6871140
## 479	1 0.6908669
## 480	1 0.7183685
## 481	1 0.7330227
## 482	1 0.8224460
## 483	1 0.8519800
## 484	1 0.6161427
## 485	1 0.8198398
## 486	1 0.7017680
## 487	1 0.7247197
## 488	1 0.6739114
## 489	1 0.8402431
## 490	1 0.6882727
## 491	1 0.7388688
## 492	1 0.7817887
## 493	1 0.8152805
## 494	1 0.6936788
## 495	1 0.6430137
## 496	1 0.7088911
## 497	1 0.8046155
## 498	1 0.7128483
## 499	1 0.7331553
## 500	1 0.8439577
## 501	2 0.8356232
## 502	2 0.8576347
## 503	2 0.9069435
## 504	2 0.7625304
## 505	2 0.7977563
## 506	2 0.7600104
## 507	2 0.8079170
## 508	2 0.7954944
## 509	2 0.8778328
## 510	2 0.8508818
## 511	2 0.7730913
## 512	2 0.7429340
## 513	2 0.7707920
## 514	2 0.7487359
## 515	2 0.8062813
## 516	2 0.8602615
## 517	2 0.7854957
## 518	2 0.8025351

## 519	2 0.9638225
## 520	2 0.8894851
## 521	2 0.9899301
## 522	2 0.9585792
## 523	2 0.8034226
## 524	2 0.8780227
## 525	2 0.8612971
## 526	2 0.9318780
## 527	2 0.8549694
## 528	2 0.7344765
## 529	2 0.9565517
## 530	2 0.8434167
## 531	2 0.8720511
## 532	2 0.8549179
## 533	2 0.7482624
## 534	2 0.7479098
## 535	2 0.8532060
## 536	2 0.7929972
## 537	2 0.8226718
## 538	2 0.7663755
## 539	2 0.7889132
## 540	2 0.8900290
## 541	2 0.8516790
## 542	2 0.8620793
## 543	2 0.8399099
## 544	2 0.7034806
## 545	2 0.8289678
## 546	2 0.6863665
## 547	2 0.6810660
## 548	2 0.7916305
## 549	2 0.8324944
## 550	2 0.8836614
## 551	2 0.7597824
## 552	2 0.7118252
## 553	2 0.8800950
## 554	2 0.8408032
## 555	2 0.9087152
## 556	2 0.7597711
## 557	2 0.8050721
## 558	2 0.7543071
## 559	2 0.8451404

## 560	2 0.8588183
## 561	2 0.8144064
## 562	2 0.8607205
## 563	2 0.8200627
## 564	2 0.7921959
## 565	2 0.8746960
## 566	2 0.8101033
## 567	2 0.7430504
## 568	2 0.8438341
## 569	2 0.9006241
## 570	2 0.6883683
## 571	2 0.8366140
## 572	2 0.7989437
## 573	2 0.8803715
## 574	2 0.7621437
## 575	2 0.8716537
## 576	2 0.7274755
## 577	2 0.7872167
## 578	2 0.8480812
## 579	2 0.7177229
## 580	2 0.7336205
## 581	2 0.8600181
## 582	2 0.8946938
## 583	2 0.7528787
## 584	2 0.8495707
## 585	2 0.7594676
## 586	2 0.7969071
## 587	2 0.8550405
## 588	2 0.8211385
## 589	2 0.8126198
## 590	2 0.7498440
## 591	2 0.8155816
## 592	2 0.7575130
## 593	2 0.8183023
## 594	2 0.8072639
## 595	2 0.8497721
## 596	2 0.8052274
## 597	2 0.6533080
## 598	2 0.7648152
## 599	2 0.7646892
## 600	2 0.8305868

## 601	2 0.8600698
## 602	2 0.8371369
## 603	2 0.8453528
## 604	2 0.8216225
## 605	2 0.7609152
## 606	2 0.8601258
## 607	2 0.6877182
## 608	2 0.8959507
## 609	2 0.7858843
## 610	2 0.7796641
## 611	2 0.8451638
## 612	2 0.8678929
## 613	2 0.9337986
## 614	2 0.7427235
## 615	2 0.8515040
## 616	2 0.7671038
## 617	2 0.8086462
## 618	2 0.8428922
## 619	2 0.8258554
## 620	2 0.8089959
## 621	2 0.6698463
## 622	2 0.8915057
## 623	2 0.8002866
## 624	2 0.7544830
## 625	2 0.8257772
## 626	2 0.7780637
## 627	2 0.7704690
## 628	2 0.7806570
## 629	2 0.7667784
## 630	2 0.8813062
## 631	2 0.8000695
## 632	2 0.8772018
## 633	2 0.8280249
## 634	2 0.8196844
## 635	2 0.8935414
## 636	2 0.9419968
## 637	2 0.8248511
## 638	2 0.6756705
## 639	2 0.7834254
## 640	2 0.8687973
## 641	2 0.8100566

## 642	2 0.7206395
## 643	2 0.7748060
## 644	2 0.8238692
## 645	2 0.8451341
## 646	2 0.9102229
## 647	2 0.8427170
## 648	2 0.8665337
## 649	2 0.8992097
## 650	2 0.7569671
## 651	2 0.8199801
## 652	2 0.8475342
## 653	2 0.7451156
## 654	2 0.7690549
## 655	2 0.8439577
## 656	2 0.8113529
## 657	2 0.8820849
## 658	2 0.7788574
## 659	2 0.7289768
## 660	2 0.8970764
## 661	2 0.8746280
## 662	2 0.8020502
## 663	2 0.8522698
## 664	2 0.7363047
## 665	2 0.8747602
## 666	2 0.8230940
## 667	2 0.8286820
## 668	2 0.7415709
## 669	2 0.7617326
## 670	2 0.7845076
## 671	2 0.8490815
## 672	2 0.8899246
## 673	2 0.7197582
## 674	2 0.6724389
## 675	2 0.8792515
## 676	2 0.8807840
## 677	2 0.8391513
## 678	2 0.7289943
## 679	2 0.9253318
## 680	2 0.8631886
## 681	2 0.8300139
## 682	2 0.8460104

## 683	2 0.8591315
## 684	2 0.8466574
## 685	2 0.8466149
## 686	2 0.8300710
## 687	2 0.8567081
## 688	2 0.8416058
## 689	2 0.7722640
## 690	2 0.8350623
## 691	2 0.8189833
## 692	2 0.7215155
## 693	2 0.7625617
## 694	2 0.8001049
## 695	2 0.7822876
## 696	2 0.8537007
## 697	2 0.9314450
## 698	2 0.7986489
## 699	2 0.8380116
## 700	2 0.8203969
## 701	2 0.8134868
## 702	2 0.9546167
## 703	2 0.8519599
## 704	2 0.8478628
## 705	2 0.7960899
## 706	2 0.8699120
## 707	2 0.7704992
## 708	2 0.8068443
## 709	2 0.8301692
## 710	2 0.7825974
## 711	2 0.8858795
## 712	2 0.8477523
## 713	2 0.8561169
## 714	2 0.7593487
## 715	2 0.8763095
## 716	2 0.9013728
## 717	2 0.8702255
## 718	2 0.7882970
## 719	2 0.7438353
## 720	2 0.8545806
## 721	2 0.8589623
## 722	2 0.8405201
## 723	2 0.7596300

## 724	2 0.8641890
## 725	2 0.6997790
## 726	2 0.7278964
## 727	2 0.7978998
## 728	2 0.7323854
## 729	2 0.7683640
## 730	2 0.7923626
## 731	2 0.8467943
## 732	2 0.8506518
## 733	2 0.9376074
## 734	2 0.6859430
## 735	2 0.9162976
## 736	2 0.7877969
## 737	2 0.7965278
## 738	2 0.7925039
## 739	2 0.9319591
## 740	2 0.8178543
## 741	2 0.8140651
## 742	2 0.8512409
## 743	2 0.8535920
## 744	2 0.7398367
## 745	2 0.7138928
## 746	2 0.7459268
## 747	2 0.8668014
## 748	2 0.7851246
## 749	2 0.7595601
## 750	2 0.8685129
## 751	9 0.8220390
## 752	9 0.8681292
## 753	9 0.9120966
## 754	9 0.7440779
## 755	9 0.7764374
## 756	9 0.7249814
## 757	9 0.8489753
## 758	9 0.9444030
## 759	9 0.8998064
## 760	9 0.7396292
## 761	9 0.7328257
## 762	9 0.6479051
## 763	9 0.8157535
## 764	9 0.7116869

## 765	9 0.7195469
## 766	9 0.9106824
## 767	9 0.6944302
## 768	9 0.7792436
## 769	9 0.9968086
## 770	9 0.8597856
## 771	9 0.9567730
## 772	9 1.1317462
## 773	9 0.8283673
## 774	9 0.8958845
## 775	9 0.9717490
## 776	9 0.8141556
## 777	9 0.8025035
## 778	9 0.7721717
## 779	9 0.8889994
## 780	9 0.8743808
## 781	9 0.9270340
## 782	9 0.7772050
## 783	9 0.8955477
## 784	9 0.7132375
## 785	9 0.8454750
## 786	9 0.8052286
## 787	9 0.6940234
## 788	9 0.9556092
## 789	9 0.7377520
## 790	9 0.9776135
## 791	9 0.8169024
## 792	9 0.8285523
## 793	9 0.8962407
## 794	9 0.6531683
## 795	9 0.8386974
## 796	9 0.6187043
## 797	9 0.6765713
## 798	9 0.7478681
## 799	9 0.9945166
## 800	9 0.8622913
## 801	9 0.6824102
## 802	9 0.7268150
## 803	9 0.8431838
## 804	9 0.8235472
## 805	9 0.9827615

## 806	9 0.8897651
## 807	9 0.7820117
## 808	9 0.6989057
## 809	9 0.7769575
## 810	9 1.0006536
## 811	9 0.7400854
## 812	9 0.9298096
## 813	9 0.8385813
## 814	9 0.8451794
## 815	9 0.8876622
## 816	9 0.7611475
## 817	9 0.7977202
## 818	9 0.9006893
## 819	9 0.8869869
## 820	9 0.7607271
## 821	9 0.8712322
## 822	9 0.6070373
## 823	9 0.9958519
## 824	9 0.8370668
## 825	9 0.7890085
## 826	9 0.5751283
## 827	9 0.8049093
## 828	9 0.9151795
## 829	9 0.7608699
## 830	9 0.7301886
## 831	9 0.8581340
## 832	9 0.8130511
## 833	9 0.7623905
## 834	9 0.7705973
## 835	9 0.7663414
## 836	9 0.6138765
## 837	9 0.7964326
## 838	9 0.8421769
## 839	9 0.8133544
## 840	9 0.6869158
## 841	9 0.8174586
## 842	9 0.6654702
## 843	9 0.8487084
## 844	9 0.6570319
## 845	9 0.7638617
## 846	9 0.8730605

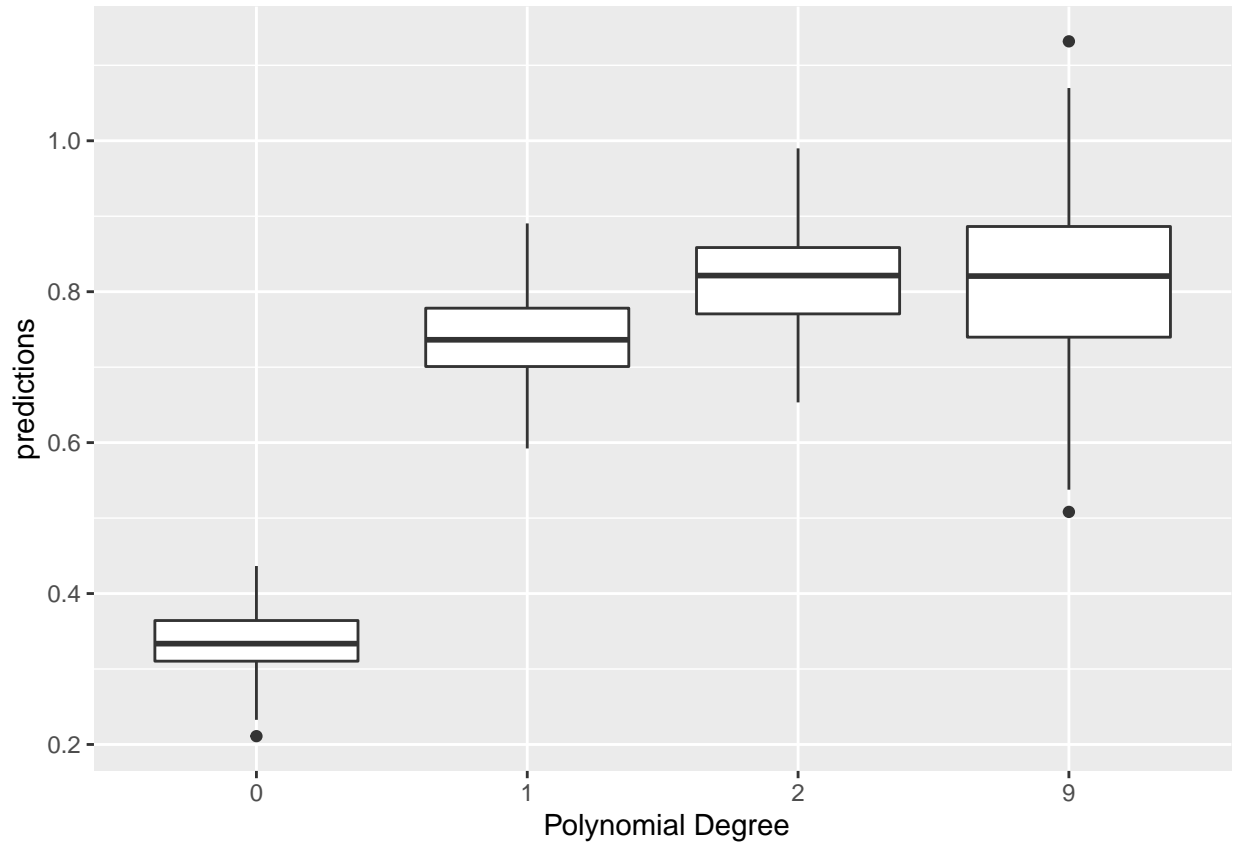
## 847	9 0.5976265
## 848	9 0.8230246
## 849	9 0.7795280
## 850	9 0.8576433
## 851	9 0.7912725
## 852	9 1.0698545
## 853	9 0.6778161
## 854	9 0.8124580
## 855	9 0.7148709
## 856	9 0.8298249
## 857	9 0.6552249
## 858	9 0.9338890
## 859	9 0.7620516
## 860	9 0.7334949
## 861	9 0.7133301
## 862	9 0.8141026
## 863	9 0.9205122
## 864	9 0.6253923
## 865	9 0.8761316
## 866	9 0.7528246
## 867	9 0.9159806
## 868	9 0.7556588
## 869	9 0.8350953
## 870	9 0.7267405
## 871	9 0.6351163
## 872	9 0.8823729
## 873	9 0.9130644
## 874	9 0.7302610
## 875	9 0.8904231
## 876	9 0.9153763
## 877	9 0.8579790
## 878	9 0.7858445
## 879	9 0.7938756
## 880	9 0.8807862
## 881	9 0.7970165
## 882	9 0.8142899
## 883	9 0.8217508
## 884	9 0.8055616
## 885	9 0.8525430
## 886	9 0.9234591
## 887	9 0.8196549

## 888	9 0.6724189
## 889	9 0.6744894
## 890	9 0.9694580
## 891	9 0.7377770
## 892	9 0.7583415
## 893	9 0.7219108
## 894	9 0.7629046
## 895	9 0.7955054
## 896	9 0.9576960
## 897	9 0.9415120
## 898	9 0.7555415
## 899	9 0.9780622
## 900	9 0.7698396
## 901	9 0.8845540
## 902	9 0.7393217
## 903	9 0.6883355
## 904	9 0.7322092
## 905	9 0.7610524
## 906	9 0.8457029
## 907	9 0.8671650
## 908	9 0.8244983
## 909	9 0.7244864
## 910	9 0.9032965
## 911	9 0.8640366
## 912	9 0.7898870
## 913	9 0.8842548
## 914	9 0.7264336
## 915	9 0.8813943
## 916	9 0.8958863
## 917	9 0.7025942
## 918	9 0.7872116
## 919	9 0.9087594
## 920	9 0.6581640
## 921	9 0.8425645
## 922	9 0.8672287
## 923	9 0.7079389
## 924	9 0.6987172
## 925	9 0.8783502
## 926	9 0.8373606
## 927	9 1.0292490
## 928	9 0.6986288

## 929	9 0.9380584
## 930	9 0.9380509
## 931	9 0.8272006
## 932	9 0.7515433
## 933	9 0.9280472
## 934	9 0.9578439
## 935	9 0.8639523
## 936	9 0.8589684
## 937	9 0.7790671
## 938	9 0.7132611
## 939	9 0.7140061
## 940	9 0.9165811
## 941	9 0.8364335
## 942	9 0.7364511
## 943	9 0.7195275
## 944	9 0.9615047
## 945	9 0.7992415
## 946	9 0.9232104
## 947	9 0.9113638
## 948	9 0.8989037
## 949	9 0.9792205
## 950	9 0.9153845
## 951	9 0.8662898
## 952	9 0.9717914
## 953	9 0.7594518
## 954	9 0.8288033
## 955	9 0.8791482
## 956	9 0.9368756
## 957	9 0.6325074
## 958	9 0.9389850
## 959	9 0.6960789
## 960	9 0.8513140
## 961	9 0.8770856
## 962	9 0.9247413
## 963	9 0.7880091
## 964	9 0.6632189
## 965	9 0.8486491
## 966	9 0.8371443
## 967	9 0.8846323
## 968	9 0.8373372
## 969	9 0.6599007

```
## 970      9 0.8704428
## 971      9 0.8395328
## 972      9 0.8938835
## 973      9 0.7898198
## 974      9 0.8192195
## 975      9 0.5082067
## 976      9 0.6478952
## 977      9 0.6802052
## 978      9 0.7593713
## 979      9 0.7381962
## 980      9 0.7903408
## 981      9 0.8092378
## 982      9 0.8362489
## 983      9 1.0083731
## 984      9 0.6421965
## 985      9 0.8945477
## 986      9 0.8122356
## 987      9 0.8879790
## 988      9 0.6809401
## 989      9 0.8489984
## 990      9 0.9121044
## 991      9 0.8706506
## 992      9 0.9683928
## 993      9 0.7608030
## 994      9 0.8720004
## 995      9 0.7414955
## 996      9 0.5375979
## 997      9 0.8710186
## 998      9 0.8433787
## 999      9 0.7020188
## 1000     9 0.9010498
```

```
ggplot(data = predictions, mapping = aes(x = degree, y = pred))+
  geom_boxplot()+
  labs(
    x = " Polynomial Degree",
    y = "predictions"
  )
```

로 위치럼 box plot이 그려진다.

이때 degree가 증가할수록 중간값이 증가하고, degree가 1,2,9일때 비슷하고, degree가 2, 9일때는 거의 유사하다.

또한, first and third quartiles (the 25th and 75th percentiles)의 차이=(IQR)가 degree가 증가할 수록 증가한다.

더하여서, degree가 0일때와 9일때 $1.5 * \text{IQR}$ from the hinge를 적용했을 때, outlying point가 존재한다.

따라서 boxplot에서 살펴볼때는 degree=9 모델보다는 degree=2 모델을 쓰는 것이 IQR이 더 적고, outlier도 없으므로, 선호된다.