PREDICTING CLIMATE PATTERNS

Team Member(s):

Balaji.M - 19BCE1512

S Lakshman Rao 19BCE1452

S.M Satya sree Narayanan 19BCE1172

S Vaibhave 19BCE1401

CODE, GRAPHS AND PREDICTIONS

**R implementation:**

1. **Temperature vs Year**

linear<-lm(ANNUAL~YEAR,data=mean1)

cubic<-lm(ANNUAL~poly(YEAR,3),data=mean1) expo<-lm(ANNUAL~poly(YEAR,2.732),data=mean1) p4<-lm(ANNUAL~poly(YEAR,4),data=mean1)

p5<-lm(ANNUAL~poly(YEAR,5),data=mean1) square<-lm(ANNUAL~poly(YEAR,2),data=mean1) loess(formula = ANNUAL ~ YEAR, data = mean1) library(MASS)

linear<-lm(ANNUAL~YEAR,data=mean1) step1.model <- stepAIC(linear, direction = "both",

trace = FALSE)

step2.model <- stepAIC(linear, direction = "forward", trace = FALSE)

step3.model <- stepAIC(linear, direction = "backward", trace = FALSE)

mean1$Xbar <- ifelse(mean1$YEAR>1990,1,0) mean1$diff <- mean1$YEAR - 1990

mean1$X <- mean1$diff\*mean1$Xbar

spline<- lm(ANNUAL ~YEAR + X, data =mean1)

## #PLOTTING THE OBTAINED MODELS

plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(P4))

plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(P5))

plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(cubic)) plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(expo))

plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(linear))

plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(loess))

plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(spline))

plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(step1.model))

plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(step2.model))

plot\_ly(mean1,x=~YEAR,y=~ANNUAL,type="scatter") %>% add\_lines(x = ~YEAR, y = fitted(step3.model))

## 2)Population Vs Year:

**Code:**

pacman::p\_load(pacman,dplyr,GGally,ggplot2,ggthemes,ggvis,httr,lubridate,plotly,rio,markdown,shin y,stringr,tidyr)

library(mgcv)

library(plotly)

a<-import("C:\\Users\\sathy\\Downloads\\Population - Sheet1 (1).csv") head(a)

plot(a$Year,a$Population,

col = "#cc0000", # Hex code for datalab.cc red pch = 19, # Use solid circles for points main = "Year vs. Population",

xlab="Year", ylab="Population(x 10^3)"

)

l<-lm(Population~Year,data=a)

j<-loess(Population~Year,data=a) summary(l)

summary(j) library(MASS) library(plotly)

linear<-lm(Population~Year,data=a) s1<- stepAIC(linear, direction = "both",

trace = FALSE)

s2<- stepAIC(linear, direction = "forward",

trace = FALSE)

s3<- stepAIC(linear, direction = "backward", trace = FALSE)

summary(s1) summary(s2) summary(s3)

b<-import("C:\\Users\\sathy\\Downloads\\dataaaa.csv") d <- lm(Population~ Year, data=a)

summary(d)

p.year <- data.frame(Year=seq(2010,2100,10)) f<-predict(d, newdata = p.year)

print("The population predicted from the model is as follows( Each value \*10^3)") print(f)

p3<-lm(Population~poly(Year,3.67),data=a) e<-lm(Population~poly(Year,2.732),data=a) p4<-lm(Population~poly(Year,4),data=a) p5<-lm(Population~poly(Year,5),data=a) p2<-lm(Population~poly(Year,2),data=a) summary(p2)

summary(p3) summary(p4) summary(p5) summary(e)

**#Plotting the dataset along with the model** plot\_ly(a,x=~Year,y=~Population,type="scatter") %>% add\_lines(x=~Year,y=fitted(e)) plot\_ly(a,x=~Year,y=~Population,type="scatter") %>% add\_lines(x=~Year,y=fitted(p2)) plot\_ly(a,x=~Year,y=~Population,type="scatter") %>% add\_lines(x=~Year,y=fitted(p3))

plot\_ly(a,x=~Year,y=~Population,type="scatter") %>% add\_lines(x=~Year,y=fitted(p4)) plot\_ly(a,x=~Year,y=~Population,type="scatter") %>% add\_lines(x=~Year,y=fitted(p5))

ggplot(data=a,aes(x=Year,y=Population))+geom\_point()

+stat\_smooth(method="lm",col="doderblue3")

+theme(panel.background=element\_rect(fill="white"),axis.line.x=element\_line(),axis.line.y=element\_l ine())+ggtitle("Regression model for Population using linear regression")

ggplot(data=a,aes(x=Year,y=Population))+geom\_point()

+stat\_smooth(method="loess",col="doderblue3")

+theme(panel.background=element\_rect(fill="white"),axis.line.x=element\_line(),axis.line.y=element\_l ine())+ggtitle("Regression model for Population using LOESS")

**3)Sea level Vs Year:** pacman::p\_load(pacman,dplyr,GGally,ggplot2,ggthemes,ggvis,httr,lubridate,plotly,rio,markdown,shin y,stringr,tidyr)

library(mgcv) library(plotly) library(MASS) **#Importing dataset**

abc=import("C:/Users/M.BALAJI/Desktop/sealevel\_final\_dataset.csv") ggpairs(data=abc,columns=1:3,title="Sea levels")

y<-as.matrix(abc[,2])

x<-as.matrix(abc[,3])

x1<-as.matrix(abc[,1])

## #Linear regression

xreg1<-lm(GMSL~Time,data=abc) summary(xreg1)

## #Loess

xreg2<-loess(y~x1,data=abc)

summary(xreg2)

## #Spline regression

spl<-gam(GMSL~s(Time),data=abc) summary(spl)

## #Poly-cubic

cbc<-lm(GMSL~poly(Time,3),data=abc) summary(cbc)

## #Poly-quadratic(square)

sqr<-lm(GMSL~poly(Time,2),data=abc) summary(sqr)

## #Poly-exponential

exp<-lm(GMSL~poly(Time,2.732),data=abc) summary(exp)

## #Poly-biquadratic(power 4)

biquad<-lm(GMSL~poly(Time,4),data=abc) summary(biquad)

## #Poly-power 5

p5<-lm(GMSL~poly(Time,5),data=abc) summary(p5)

## #Stepwise-forward and backward

step1<-stepAIC(xreg1,direction="both",trace=FALSE) summary(step1)

## #Stepwise-forward

step2<-stepAIC(xreg1,direction="forward",trace=FALSE) summary(step2)

## #Stepwise-backward

step3<-stepAIC(xreg1,direction="backward",trace=FALSE) summary(step3)

## #PLOTTING THE OBTAINED MODELS

ggplot(data=abc,aes(x=Time,y=GMSL))+geom\_point()+stat\_smooth(method="lm",col="dogerblue3")

+theme(panel.background=element\_rect(fill="white"),axis.line.x=element\_line(),axis.line.y = element\_line())+ggtitle("Linear model fitted to Data") plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(xreg1)) ggplot(data=abc,aes(x=Time,y=GMSL))+geom\_point()

+stat\_smooth(method="loess",col="dogerblue3")

+theme(panel.background=element\_rect(fill="white"),axis.line.x=element\_line(),axis.line.y = element\_line())+ggtitle("Loess model fitted to Data") plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(xreg2)) ggplot(data=abc,aes(x=Time,y=GMSL))+geom\_point()

+stat\_smooth(method="gam",col="dogerblue3")

+theme(panel.background=element\_rect(fill="white"),axis.line.x=element\_line(),axis.line.y = element\_line())+ggtitle("spline model fitted to Data") plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(spl)) plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(cbc)) plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(sqr)) plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(exp)) plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(biquad)) plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(p5)) plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(step1)) plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(step2)) plot\_ly(abc,x=~Time,y=~GMSL,type="scatter") %>% add\_lines(x=~Time,y=fitted(step3)) **#PREDICTION**

new\_year<-data.frame(Time=seq(2010,2100,10)) predict(spl,newdata=new\_year)

## 4)CO2 levels Vs Year:

library(datasets) # Load base packages manually library(ggplot2)

library(GGally) library(mgcv)

library(MASS)

**# Install pacman ("package manager") if needed** if (!require("pacman")) install.packages("pacman") pacman::p\_load(pacman, caret, lars, tidyverse) pacman::p\_load(pacman, rio)

## # Excel XLSX

rio\_xlsx <- import("C:\\Users\\Lakshman\\Desktop\\CO2.xlsx") head(rio\_xlsx)

str(rio\_xlsx) plot(rio\_xlsx$Year,rio\_xlsx$Total)

ggpairs(data=rio\_xlsx,columns=1:8,title="CO2 levels") x<-as.matrix(rio\_xlsx[2])

x

y<-rio\_xlsx[,2] y

**##quadratic**

m2<-lm(Total~poly(Year,2,raw=T),data=rio\_xlsx) summary(m2)

**##Exponential**

m21<-lm(Total~poly(Year,2.732,raw=T),data=rio\_xlsx) summary(m21)

**##Cubic**

m3<-lm(Total~poly(Year,3,raw=T),data=rio\_xlsx) summary(m3)

**##To the power 4**

m4<-lm(Total~poly(Year,4,raw=TRUE),data=rio\_xlsx) summary(m4)

**##To the power 5**

m5<-lm(Total~poly(Year,5,raw=TRUE),data=rio\_xlsx) summary(m5)

**##loess**

seg1<-loess(y~x,data=rio\_xlsx) summary(seg1)

**##spline regression**

model<-gam(Total~s(Year),data =rio\_xlsx) summary(model)

**##Stepwise**

**##1.Both**

rem<-lm(Total~Year,data=rio\_xlsx)

step1.model<-stepAIC(rem,direction="both",trace=F)

summary(step1.model)

**##2.Forward**

step2.model<-stepAIC(rem,direction="forward",trace=F) summary(step2.model)

**##3.Backward**

step3.model<-stepAIC(rem,direction="backward",trace=F) summary(step3.model)

## #PLOTTING THE OBTAINED MODELS

plot\_ly(rio\_xlsx,x=~Year,y=~Total,type="scatter") %>% add\_lines(x= ~Year, y= fitted(m2)) plot\_ly(rio\_xlsx,x=~Year,y=~Total,type="scatter") %>% add\_lines(x= ~Year, y= fitted(m21)) plot\_ly(rio\_xlsx,x=~Year,y=~Total,type="scatter") %>% add\_lines(x= ~Year, y= fitted(m3)) plot\_ly(rio\_xlsx,x=~Year,y=~Total,type="scatter") %>% add\_lines(x= ~Year, y= fitted(m4)) plot\_ly(rio\_xlsx,x=~Year,y=~Total,type="scatter") %>% add\_lines(x= ~Year, y= fitted(m5))

ggplot(data=rio\_xlsx,aes(x=Year,y=Total))+geom\_point()

+stat\_smooth(method="loess",col="dodgerblue3")

+theme(panel.background=element\_rect(fill="white"),axis.line.x=element\_line(),axis.line.y=elem ent\_line())

ggplot(rio\_xlsx,aes(Year,Total))+geom\_point()+stat\_smooth(method=gam,formula=y~s(x))

plot\_ly(rio\_xlsx,x=~Year,y=~Total,type="scatter")%>%add\_lines(x=~Year,y=fitted(step1.model))

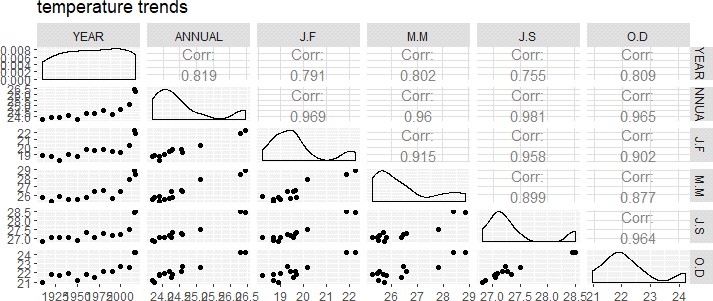
plot\_ly(rio\_xlsx,x=~Year,y=~Total,type="scatter")%>%add\_lines(x=~Year,y=fitted(step2.model)) plot\_ly(rio\_xlsx,x=~Year,y=~Total,type="scatter")%>%add\_lines(x=~Year,y=fitted(step3.model)) **#PREDICTION**

predict(m4,data.frame(Year= seq(2010,2100,10)))

# Correlation using ggpairs:

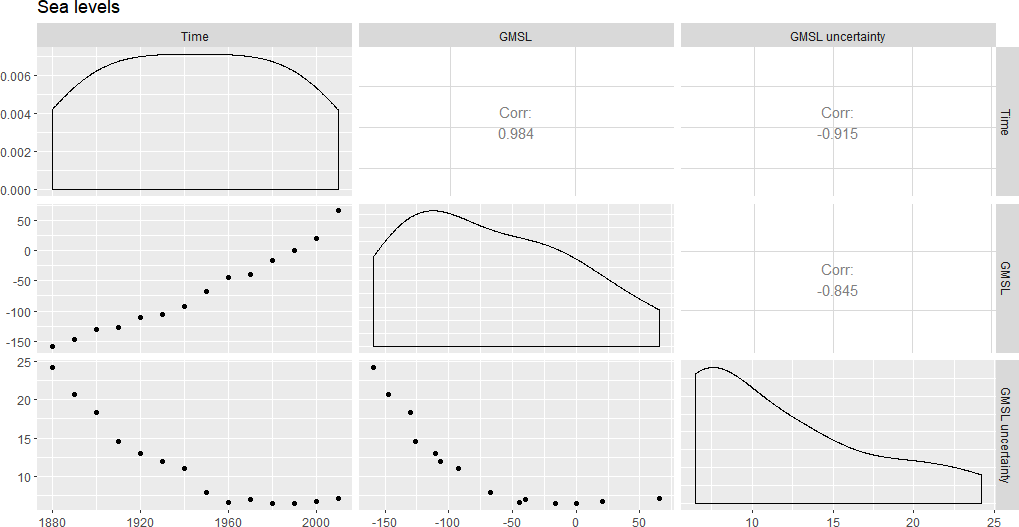
## Temperature Vs Year:

ggpairs(data=mean1, columns=1:6, title="temperature trends")



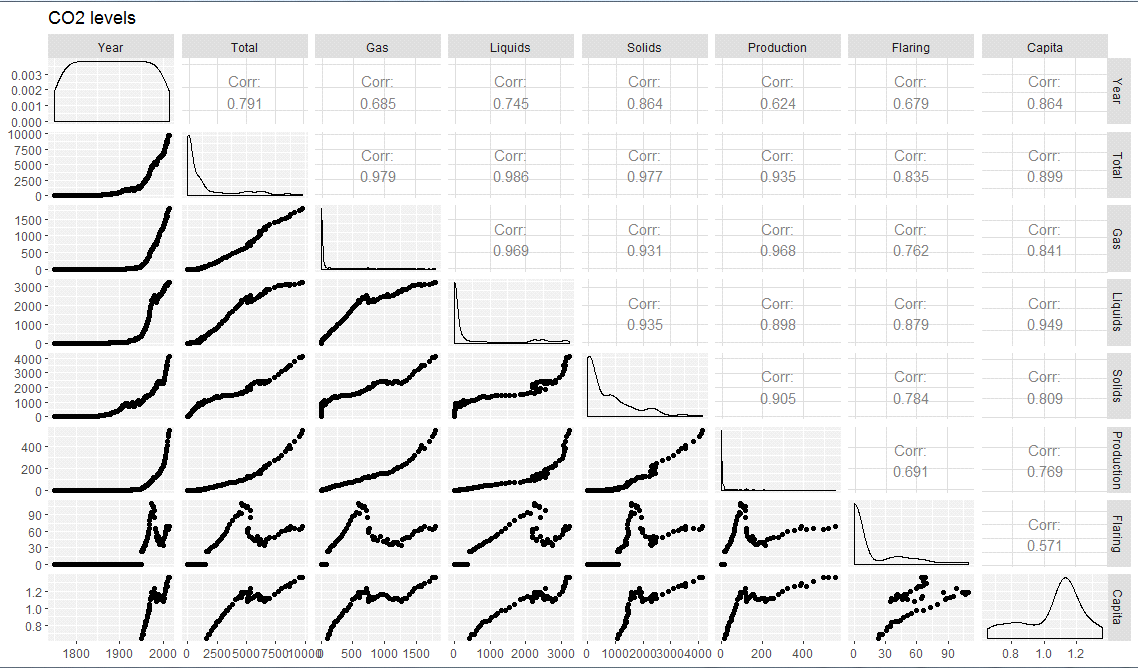
We can see that there is a 82% correlation between the annual temperature and the year we can predict the temperature given the year with a decent accuracy.

## Sea level Vs Year:

ggpairs(data=abc,columns=1:3,title="Sea levels")

Since there is a 98% correlation between Time and GSML, we can predict sea level using given year (Time) very accurately.

1. **CO2 levels Vs Year:**ggpairs(data=rio\_xlsx , columns=1:8,title="CO2 levels")



We can clearly that there is an 80% correlation between the annual CO2 emissions and the year, thus we can predict the temperature given the year with decent accuracy.

# Outputs of all the models:

## Temperature Vs Year: i)Spline Regression:

Residuals:

Min 1Q Median 3Q Max

-0.52058 -0.05522 0.03142 0.14994 0.37987

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 13.633582 7.036594 1.938 0.081414 .

YEAR 0.005336 0.003605 1.480 0.169652

X 0.064589 0.012451 5.187 0.000409 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2926 on 10 degrees of freedom Multiple R-squared: 0.9107, Adjusted R-squared: 0.8729

F-statistic: 51 on 2 and 10 DF, p-value: 5.672e-06



## LOESS:

Number of Observations: 13 Equivalent Number of Parameters: 4.67 Residual Standard Error: 0.2319

Trace of smoother matrix: 5.15 (exact)

Control settings:

span : 0.75

degree : 2

family : gaussian

surface : interpolate cell = 0.2 normalize: TRUE

parametric: FALSE drop.square: FALSE



## Step Wise Linear Regression:

FRONT

Call:

lm(formula = ANNUAL ~ YEAR, data = mean1)

Residuals:

Min 1Q Median 3Q Max

-0.76269 -0.31666 -0.08865 0.32339 0.96249

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -14.431528 8.241007 -1.751 0.107711

YEAR 0.019801 0.004185 4.731 0.000618 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.536 on 11 degrees of freedom Multiple R-squared: 0.6705, Adjusted R-squared: 0.6405

F-statistic: 22.38 on 1 and 11 DF, p-value: 0.0006183

BACKWARD

Call:

lm(formula = ANNUAL ~ YEAR, data = mean1)

Residuals:

Min 1Q Median 3Q Max

-0.76269 -0.31666 -0.08865 0.32339 0.96249

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -14.431528 8.241007 -1.751 0.107711

YEAR 0.019801 0.004185 4.731 0.000618 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.536 on 11 degrees of freedom Multiple R-squared: 0.6705, Adjusted R-squared: 0.6405

F-statistic: 22.38 on 1 and 11 DF, p-value: 0.0006183

BOTH

Call:

lm(formula = ANNUAL ~ YEAR, data = mean1)

Residuals:

Min 1Q Median 3Q Max

-0.76269 -0.31666 -0.08865 0.32339 0.96249

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -14.431528 8.241007 -1.751 0.107711

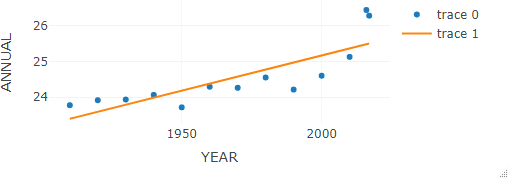
YEAR 0.019801 0.004185 4.731 0.000618 \*\*\*

---

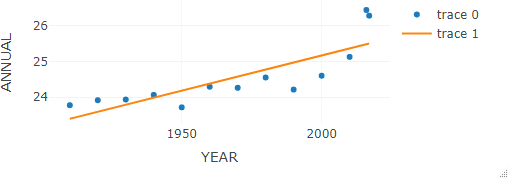
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.536 on 11 degrees of freedom Multiple R-squared: 0.6705, Adjusted R-squared: 0.6405

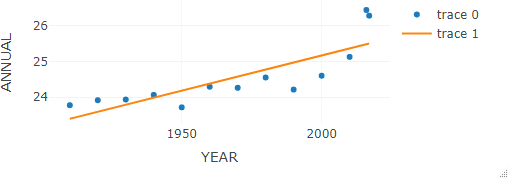
F-statistic: 22.38 on 1 and 11 DF, p-value: 0.0006183 Forward:



Backward:



Both:



## v)Power 4 and Power 5:

POWER 4

Residuals:

Min 1Q Median 3Q Max

-0.43578 -0.09607 0.03557 0.08773 0.33018

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 24.5508 0.0702 349.711 < 2e-16 \*\*\*

poly(YEAR, 4)1 2.5360 0.2531 10.019 8.37e-06 \*\*\*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| poly(YEAR, 4)2 | 1.2888 | 0.2531 | 5.092 | 0.00094 \*\*\* |
| poly(YEAR, 4)3 | 0.8153 | 0.2531 | 3.221 | 0.01222 \* |
| poly(YEAR, 4)4 | 0.5678 | 0.2531 | 2.243 | 0.05515 . |
| --- |  |  |  |  |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2531 on 8 degrees of freedom Multiple R-squared: 0.9066, Adjusted R-squared: 0.8198

F-statistic: 35.43 on 4 and 8 DF, p-value: 3.903e-05

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ POWER 5

Residuals:

Min 1Q Median 3Q Max

-0.26905 -0.07879 0.01799 0.13046 0.22189

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 24.5508 0.0554 443.119 < 2e-16 \*\*\*

poly(YEAR, 5)1 2.5360 0.1998 12.695 4.35e-06 \*\*\*

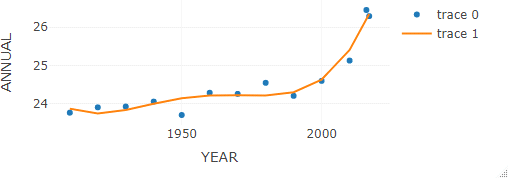
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| poly(YEAR, 5)2 | 1.2888 | 0.1998 | 6.451 | 0.00035 \*\*\* |
| poly(YEAR, 5)3 | 0.8153 | 0.1998 | 4.081 | 0.00468 \*\* |
| poly(YEAR, 5)4 | 0.5678 | 0.1998 | 2.842 | 0.02496 \* |
| poly(YEAR, 5)5 | 0.4829 | 0.1998 | 2.418 | 0.04626 \* |
| --- |  |  |  |  |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

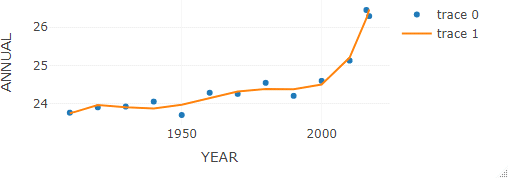
Residual standard error: 0.1998 on 7 degrees of freedom Multiple R-squared: 0.8709, Adjusted R-squared: 0.8501

F-statistic: 46.67 on 5 and 7 DF, p-value: 3.161e-05

Power 4:



Power 5:



## Cubic And Square:

CUBIC:

Call:

lm(formula = ANNUAL ~ poly(YEAR, 3), data = mean1)

Residuals:

Min 1Q Median 3Q Max

-0.42128 -0.21361 -0.01731 0.20893 0.38332

Coefficients:

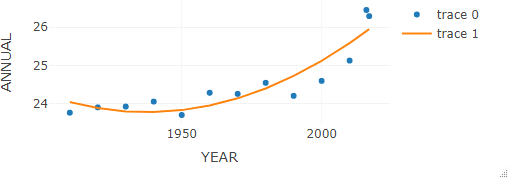
Estimate Std. Error t value Pr(>|t|) (Intercept) 24.55077 0.08448 290.618 < 2e-16 \*\*\*

|  |  |  |  |
| --- | --- | --- | --- |
| poly(YEAR, 3)1 | 2.53599 | 0.30459 | 8.326 1.61e-05 \*\*\* |
| poly(YEAR, 3)2 | 1.28877 | 0.30459 | 4.231 0.0022 \*\* |
| poly(YEAR, 3)3 | 0.81531 | 0.30459 | 2.677 0.0253 \* |
| --- |  |  |  |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Residual standard error: 0.3046 on 9 degrees of freedom

Multiple R-squared: 0.913, Adjusted R-squared: 0.8839

F-statistic: 31.46 on 3 and 9 DF, p-value: 4.224e-05



~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ SQUARE:

Call:

lm(formula = ANNUAL ~ poly(YEAR, 2), data = mean1)

Residuals:

Min 1Q Median 3Q Max

-0.5234 -0.2766 0.1146 0.2736 0.5518

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 24.5508 0.1074 228.578 < 2e-16 \*\*\*

poly(YEAR, 2)1 2.5360 0.3873 6.549 6.48e-05 \*\*\*

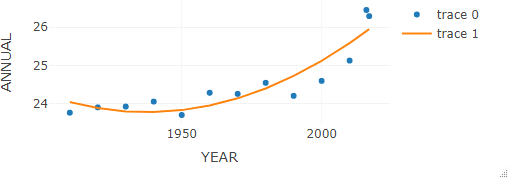
poly(YEAR, 2)2 1.2888 0.3873 3.328 0.00764 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3873 on 10 degrees of freedom Multiple R-squared: 0.8436, Adjusted R-squared: 0.8124

F-statistic: 26.98 on 2 and 10 DF, p-value: 9.343e-05



## Exponential:

EXPONENTIAL

Call:

lm(formula = ANNUAL ~ poly(YEAR, 2.732), data = mean1)

Residuals:

Min 1Q Median 3Q Max

-0.5234 -0.2766 0.1146 0.2736 0.5518

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 24.5508 0.1074 228.578 < 2e-16 \*\*\*

poly(YEAR, 2.732)1 2.5360 0.3873 6.549 6.48e-05 \*\*\*

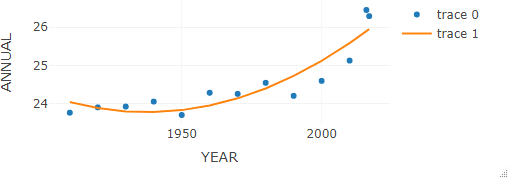
poly(YEAR, 2.732)2 1.2888 0.3873 3.328 0.00764 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3873 on 10 degrees of freedom Multiple R-squared: 0.8436, Adjusted R-squared: 0.8124

F-statistic: 26.98 on 2 and 10 DF, p-value: 9.343e-05



## Prediction:

>fit2 <- lm(ANNUAL ~ poly(YEAR,3.0), data=mean1)

>summary(fit2)

>new.year <- data.frame( YEAR=c(2020)

)

>predict(fit2, newdata = new.year)

1 2 3 4 5 6 7 8

25.55128 26.51699 27.81390 29.48951 31.59129 34.16670 37.26323 40.92836

9 10

45.20956 50.15431

## Population Vs Year:

1. **Linear Regression:**

* summary(l)

Call:

lm(formula = Population ~ Year, data = a)

Residuals:

Min 1Q Median 3Q Max

-115821 -93643 7378 49218 280189

Coefficients:

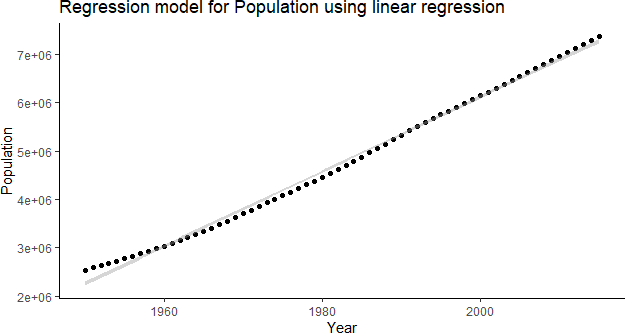
Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.484e+08 1.234e+06 -120.3 <2e-16 \*\*\* Year 7.725e+04 6.223e+02 124.1 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 96310 on 64 degrees of freedom Multiple R-squared: 0.9959, Adjusted R-squared: 0.9958 F-statistic: 1.541e+04 on 1 and 64 DF, p-value: < 2.2e-16



## LOESS:

* summary(j) Call:

loess(formula = Population ~ Year, data = a)

Number of Observations: 66 Equivalent Number of Parameters: 4.4 Residual Standard Error: 12150

Trace of smoother matrix: 4.81 (exact)

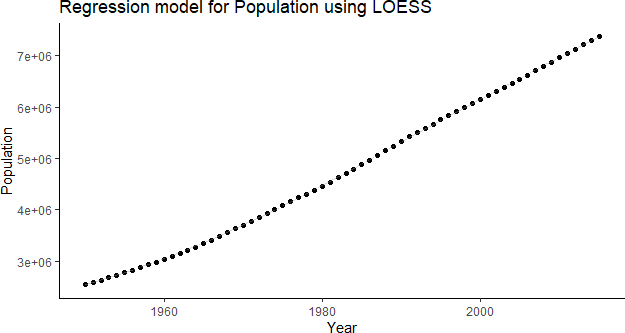
Control settings:

span : 0.75

degree : 2 family : gaussian

surface : interpolate cell = 0.2 normalize: TRUE

parametric: FALSE drop.square: FALSE



## Step Wise Linear Regression:

1. Direction:Both

* summary(s1)

Call:

lm(formula = Population ~ Year, data = a)

Residuals:

Min 1Q Median 3Q Max

-115821 -93643 7378 49218 280189

Coefficients:

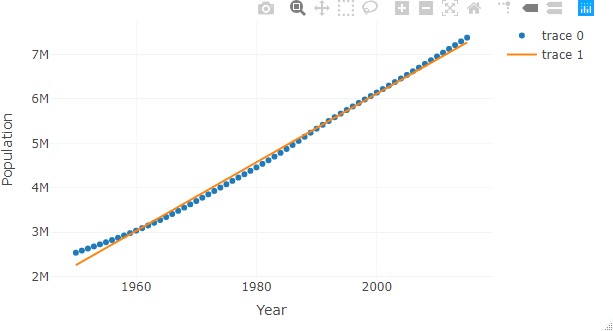
Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.484e+08 1.234e+06 -120.3 <2e-16 \*\*\* Year 7.725e+04 6.223e+02 124.1 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 96310 on 64 degrees of freedom Multiple R-squared: 0.9959, Adjusted R-squared: 0.9958 F-statistic: 1.541e+04 on 1 and 64 DF, p-value: < 2.2e-16



1. Direction:Forward

* summary(s2)

Call:

lm(formula = Population ~ Year, data = a)

Residuals:

Min 1Q Median 3Q Max

-115821 -93643 7378 49218 280189

Coefficients:

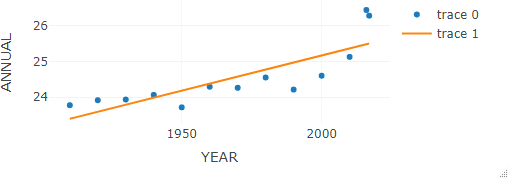
Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.484e+08 1.234e+06 -120.3 <2e-16 \*\*\* Year 7.725e+04 6.223e+02 124.1 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 96310 on 64 degrees of freedom Multiple R-squared: 0.9959, Adjusted R-squared: 0.9958 F-statistic: 1.541e+04 on 1 and 64 DF, p-value: < 2.2e-16



1. Direction:Backward

* summary(s3)

Call:

lm(formula = Population ~ Year, data = a)

Residuals:

Min 1Q Median 3Q Max

-115821 -93643 7378 49218 280189

Coefficients:

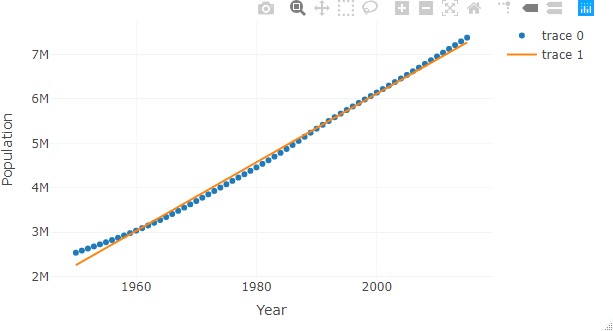
Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.484e+08 1.234e+06 -120.3 <2e-16 \*\*\* Year 7.725e+04 6.223e+02 124.1 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 96310 on 64 degrees of freedom Multiple R-squared: 0.9959, Adjusted R-squared: 0.9958 F-statistic: 1.541e+04 on 1 and 64 DF, p-value: < 2.2e-16



## Power 4 and Power 5:

Call:

lm(formula = Population ~ poly(Year, 4), data = a)

Residuals:

Min 1Q Median 3Q Max

-22112.8 -10283.1 -109.8 7691.4 27149.7

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 4766790 1672 2850.051 < 2e-16 \*\*\*

poly(Year, 4)1 11956045 13588 879.918 < 2e-16 \*\*\*

|  |  |  |  |
| --- | --- | --- | --- |
| poly(Year, 4)2 | 669588 | 13588 | 49.279 < 2e-16 \*\*\* |
| poly(Year, 4)3 | -362096 | 13588 | -26.649 < 2e-16 \*\*\* |
| poly(Year, 4)4 | 54396 | 13588 | 4.003 0.000172 \*\*\* |
| --- |  |  |  |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13590 on 61 degrees of freedom Multiple R-squared: 0.9399, Adjusted R-squared: 0.9299 F-statistic: 1.944e+05 on 4 and 61 DF, p-value: < 2.2e-16

Call:

lm(formula = Population ~ poly(Year, 4), data = a)

Residuals:

Min 1Q Median 3Q Max

-22112.8 -10283.1 -109.8 7691.4 27149.7

Coefficients:

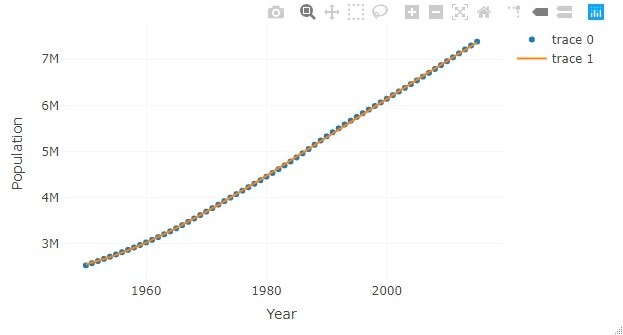
Estimate Std. Error t value Pr(>|t|) (Intercept) 4766790 1672 2850.051 < 2e-16 \*\*\*

poly(Year, 4)1 11956045 13588 879.918 < 2e-16 \*\*\*

|  |  |  |  |
| --- | --- | --- | --- |
| poly(Year, 4)2 | 669588 | 13588 | 49.279 < 2e-16 \*\*\* |
| poly(Year, 4)3 | -362096 | 13588 | -26.649 < 2e-16 \*\*\* |
| poly(Year, 4)4 | 54396 | 13588 | 4.003 0.000172 \*\*\* |
| --- |  |  |  |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13590 on 61 degrees of freedom Multiple R-squared: 0.9399, Adjusted R-squared: 0.9399 F-statistic: 1.944e+05 on 4 and 61 DF, p-value: < 2.2e-16



Power 5:

Call:

lm(formula = Population ~ poly(Year, 5), data = a)

Residuals:

Min 1Q Median 3Q Max

-22760.4 -5944.2 -42.7 7141.2 17555.5

Coefficients:

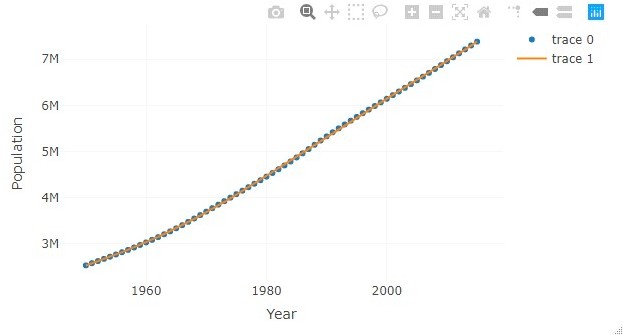
Estimate Std. Error t value Pr(>|t|) (Intercept) 4766790 1299 3669.889 < 2e-16 \*\*\*

poly(Year, 5)1 11956045 10552 1133.033 < 2e-16 \*\*\*

|  |  |  |  |
| --- | --- | --- | --- |
| poly(Year, 5)2 | 669588 | 10552 | 63.455 < 2e-16 \*\*\* |
| poly(Year, 5)3 | -362096 | 10552 | -34.315 < 2e-16 \*\*\* |
| poly(Year, 5)4 | 54396 | 10552 | 5.155 2.99e-06 \*\*\* |
| poly(Year, 5)5 | 67684 | 10552 | 6.414 2.47e-08 \*\*\* |
| --- |  |  |  |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10550 on 60 degrees of freedom Multiple R-squared: 0.9235, Adjusted R-squared: 0.9135 F-statistic: 2.578e+05 on 5 and 60 DF, p-value: < 2.2e-16



## Cubic and Square:

Call:

lm(formula = Population ~ poly(Year, 3.67), data = a)

Residuals:

Min 1Q Median 3Q Max

-26357 -10505 -1109 7501 37080

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 4766790 1864 2556.98 <2e-16 \*\*\*

poly(Year, 3.67)1 11956045 15145 789.44 <2e-16 \*\*\*

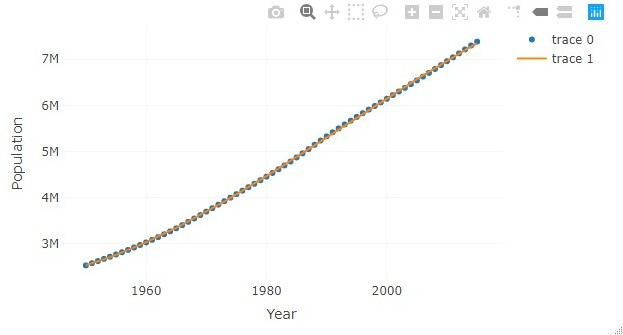
poly(Year, 3.67)2 669588 15145 44.21 <2e-16 \*\*\*

poly(Year, 3.67)3 -362096 15145 -23.91 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15150 on 62 degrees of freedom Multiple R-squared: 0.9499, Adjusted R-squared: 0.9239 F-statistic: 2.086e+05 on 3 and 62 DF, p-value: < 2.2e-16



Square:

Call:

lm(formula = Population ~ poly(Year, 2), data = a)

Residuals:

Min 1Q Median 3Q Max

-70591 -38728 -12553 41205 104082

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 4766790 5912 806.28 <2e-16 \*\*\*

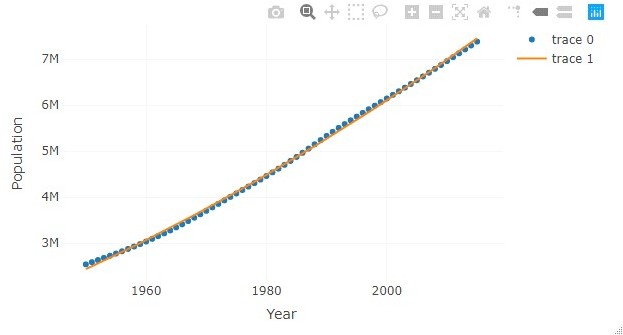
poly(Year, 2)1 11956045 48030 248.93 <2e-16 \*\*\*

poly(Year, 2)2 669588 48030 13.94 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 48030 on 63 degrees of freedom Multiple R-squared: 0.949, Adjusted R-squared: 0.949 F-statistic: 3.108e+04 on 2 and 63 DF, p-value: < 2.2e-16



## Exponential:

Call:

lm(formula = Population ~ poly(Year, 2.732), data = a)

Residuals:

Min 1Q Median 3Q Max

-70591 -38728 -12553 41205 104082

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 4766790 5912 806.28 <2e-16 \*\*\*

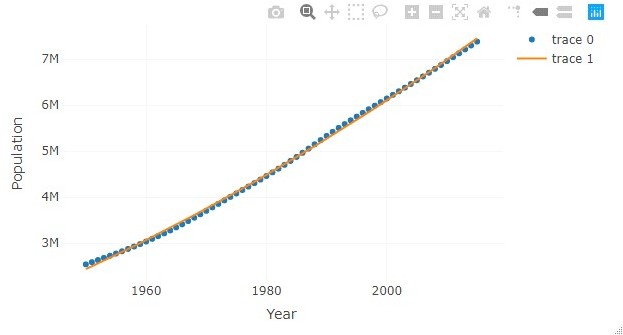
poly(Year, 2.732)1 11956045 48030 248.93 <2e-16 \*\*\*

poly(Year, 2.732)2 669588 48030 13.94 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 48030 on 63 degrees of freedom Multiple R-squared: 0.9439, Adjusted R-squared: 0.939 F-statistic: 3.108e+04 on 2 and 63 DF, p-value: < 2.2e-16



## Prediction:

* p.year <- data.frame(Year=seq(2010,2100,10))
* f<-predict(d, newdata = p.year)
* print("The population predicted from the model is as follows( Each value \*10^3)")

[1] "The population predicted from the model is as follows( Each value \*10^3)"

* print(f)

1 2 3 4 5 6 7 8 9

6891232 7663756 8436280 9208804 9981329 10753853 11526377 12298901 13071426

10

13843950

## Sea level Vs Year:

1. **Spline Regression:**

* summary(spl)

Family: gaussian Link function: identity

Formula:

GMSL ~ s(Time)

Parametric coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -68.093 1.309 -52.02 3.8e-10 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

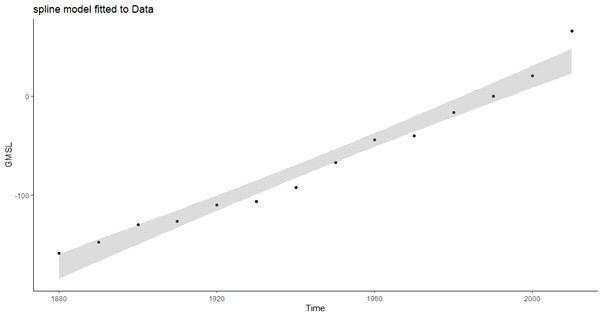
Approximate significance of smooth terms: edf Ref.df F p-value

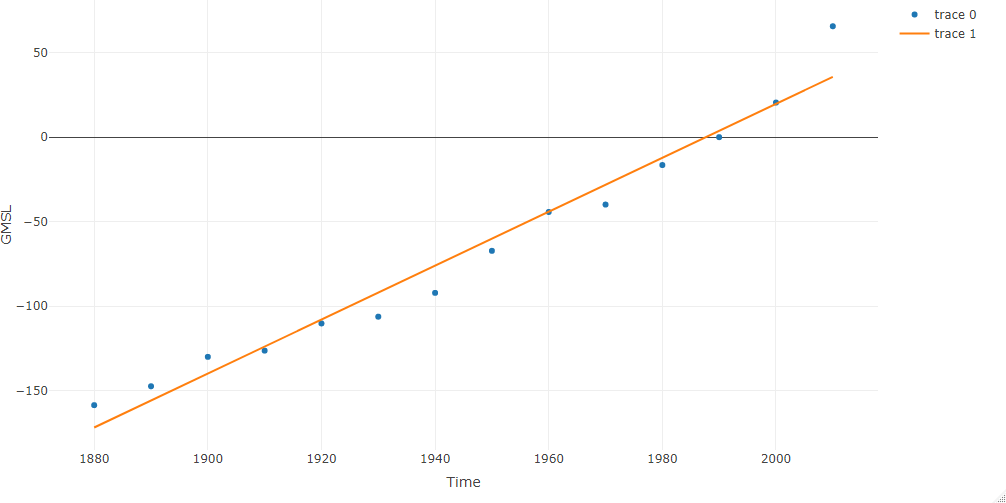
s(Time) 6.158 7.314 340.4 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) = 0.995 Deviance explained = 99.7% GCV = 49.091 Scale est. = 23.993 n = 14





## LOESS:

* summary(gam(y~x1,data=abc))

Family: gaussian Link function: identity

Formula:

y ~ x1

Parametric coefficients:

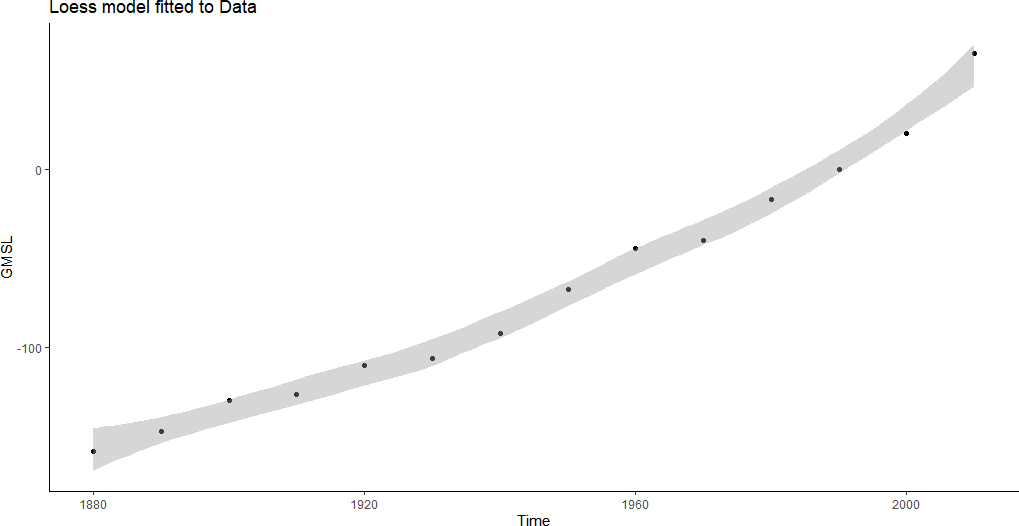
Estimate Std. Error t value Pr(>|t|)

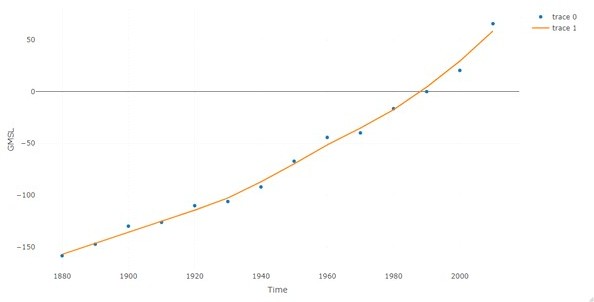
(Intercept) -3.175e+03 1.639e+02 -19.37 2.03e-10 \*\*\* x1 1.597e+00 8.426e-02 18.96 2.60e-10 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) = 0.965 Deviance explained = 96.8% GCV = 188.46 Scale est. = 161.54 n = 14





## Step Wise linear regression: 1)Forward and Backward:

* + summary(step1)

Call:

lm(formula = GMSL ~ Time, data = abc)

Residuals:

Min 1Q Median 3Q Max

-16.120 -6.475 -2.334 6.522 29.960

Coefficients:

Estimate Std. Error t value Pr(>|t|)

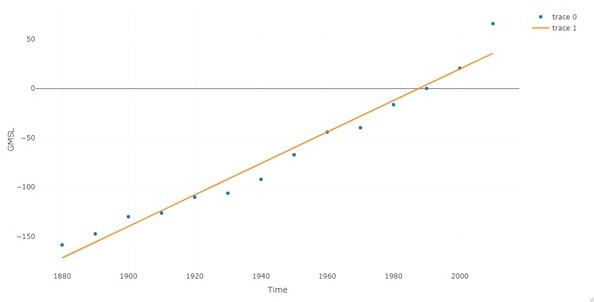
(Intercept) -3.175e+03 1.639e+02 -19.37 2.03e-10 \*\*\* Time 1.597e+00 8.426e-02 18.96 2.60e-10 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.71 on 12 degrees of freedom Multiple R-squared: 0.9677, Adjusted R-squared: 0.965

F-statistic: 359.4 on 1 and 12 DF, p-value: 2.604e-10



## Forward:

* summary(step2)

Call:

lm(formula = GMSL ~ Time, data = abc)

Residuals:

Min 1Q Median 3Q Max

-16.120 -6.475 -2.334 6.522 29.960

Coefficients:

Estimate Std. Error t value Pr(>|t|)

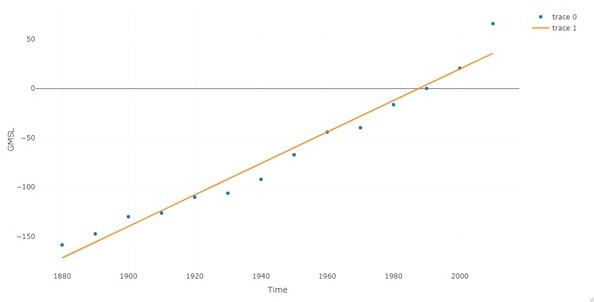
(Intercept) -3.175e+03 1.639e+02 -19.37 2.03e-10 \*\*\* Time 1.597e+00 8.426e-02 18.96 2.60e-10 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.71 on 12 degrees of freedom Multiple R-squared: 0.9677, Adjusted R-squared: 0.965

F-statistic: 359.4 on 1 and 12 DF, p-value: 2.604e-10



## Backward:

* + summary(step3)

Call:

lm(formula = GMSL ~ Time, data = abc)

Residuals:

Min 1Q Median 3Q Max

-16.120 -6.475 -2.334 6.522 29.960

Coefficients:

Estimate Std. Error t value Pr(>|t|)

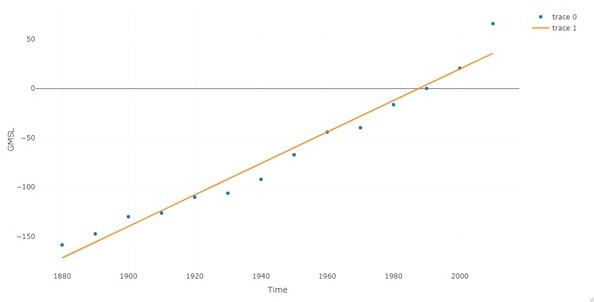
(Intercept) -3.175e+03 1.639e+02 -19.37 2.03e-10 \*\*\* Time 1.597e+00 8.426e-02 18.96 2.60e-10 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.71 on 12 degrees of freedom Multiple R-squared: 0.9677, Adjusted R-squared: 0.965

F-statistic: 359.4 on 1 and 12 DF, p-value: 2.604e-10



## Power 4 and Power 5:

Power 4:

* summary(biquad)

Call:

lm(formula = GMSL ~ poly(Time, 4), data = abc)

Residuals:

Min 1Q Median 3Q Max

-8.1678 -2.4185 -0.5767 3.0363 10.7520

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -68.093 1.801 -37.815 3.14e-11 \*\*\*

poly(Time, 4)1 240.942 6.738 35.761 5.17e-11 \*\*\*

poly(Time, 4)2 37.389 6.738 5.549 0.000357 \*\*\*

poly(Time, 4)3 10.072 6.738 1.495 0.169149

poly(Time, 4)4 5.526 6.738 0.820 0.433321

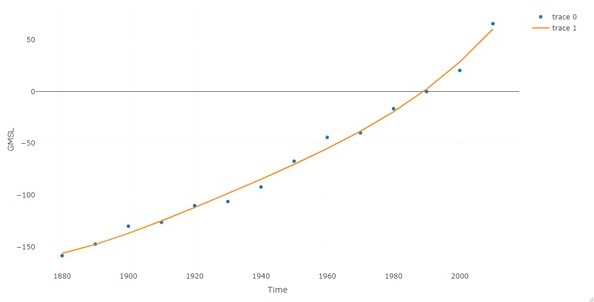
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.738 on 9 degrees of freedom

Multiple R-squared: 0.9932, Adjusted R-squared: 0.9902

F-statistic: 328.1 on 4 and 9 DF, p-value: 9.708e-10



Power 5:

* summary(p5) Call:

lm(formula = GMSL ~ poly(Time, 5), data = abc)

Residuals:

Min 1Q Median 3Q Max

-5.680 -3.254 1.162 2.100 6.530

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -68.093 1.367 -49.824 2.92e-11 \*\*\*

poly(Time, 5)1 240.942 5.114 47.118 4.55e-11 \*\*\*

poly(Time, 5)2 37.389 5.114 7.312 8.29e-05 \*\*\*

poly(Time, 5)3 10.072 5.114 1.970 0.0844 .

poly(Time, 5)4 5.526 5.114 1.081 0.3114

poly(Time, 5)5 14.120 5.114 2.761 0.0246 \*

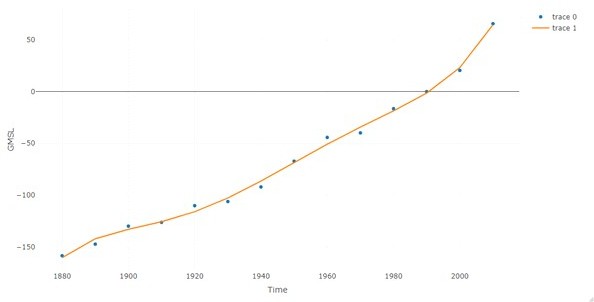
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.114 on 8 degrees of freedom

Multiple R-squared: 0.9965, Adjusted R-squared: 0.9943

F-statistic: 457.2 on 5 and 8 DF, p-value: 1.328e-09



## Exponential:

* summary(exp)

Call:

lm(formula = GMSL ~ poly(Time, 2.732), data = abc)

Residuals:

Min 1Q Median 3Q Max

-8.966 -4.808 -1.397 4.466 11.946

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -68.093 1.873 -36.346 8.24e-13 \*\*\*

poly(Time, 2.732)1 240.942 7.010 34.371 1.52e-12 \*\*\*

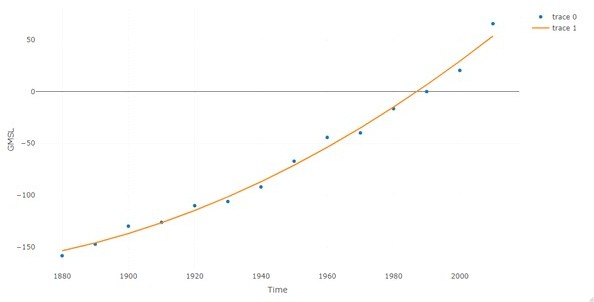
poly(Time, 2.732)2 37.389 7.010 5.334 0.00024 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.01 on 11 degrees of freedom Multiple R-squared: 0.991, Adjusted R-squared: 0.9894

F-statistic: 604.9 on 2 and 11 DF, p-value: 5.637e-12



## Cubic and Square:

Cubic:

* summary(cbc)

Call:

lm(formula = GMSL ~ poly(Time, 3), data = abc)

Residuals:

Min 1Q Median 3Q Max

-9.3210 -4.0211 -0.5315 3.9049 11.6955

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -68.093 1.771 -38.450 3.38e-12 \*\*\*

poly(Time, 3)1 240.942 6.626 36.361 5.88e-12 \*\*\*

poly(Time, 3)2 37.389 6.626 5.642 0.000215 \*\*\*

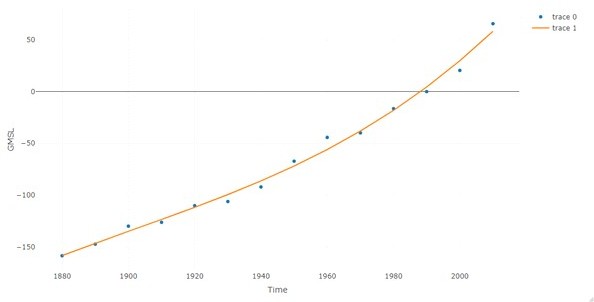
poly(Time, 3)3 10.072 6.626 1.520 0.159477

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.626 on 10 degrees of freedom Multiple R-squared: 0.9927, Adjusted R-squared: 0.9905

F-statistic: 452.1 on 3 and 10 DF, p-value: 5.668e-11



Square:

* summary(sqr)

Call:

lm(formula = GMSL ~ poly(Time, 2), data = abc)

Residuals:

Min 1Q Median 3Q Max

-8.966 -4.808 -1.397 4.466 11.946

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -68.093 1.873 -36.346 8.24e-13 \*\*\*

poly(Time, 2)1 240.942 7.010 34.371 1.52e-12 \*\*\*

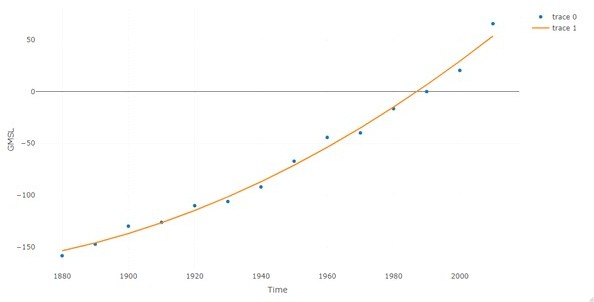
poly(Time, 2)2 37.389 7.010 5.334 0.00024 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.01 on 11 degrees of freedom Multiple R-squared: 0.991, Adjusted R-squared: 0.9894

F-statistic: 604.9 on 2 and 11 DF, p-value: 5.637e-12



## Linear regression:

* summary(xreg1)

Call:

lm(formula = GMSL ~ Time, data = abc)

Residuals:

Min 1Q Median 3Q Max

-16.120 -6.475 -2.334 6.522 29.960

Coefficients:

Estimate Std. Error t value Pr(>|t|)

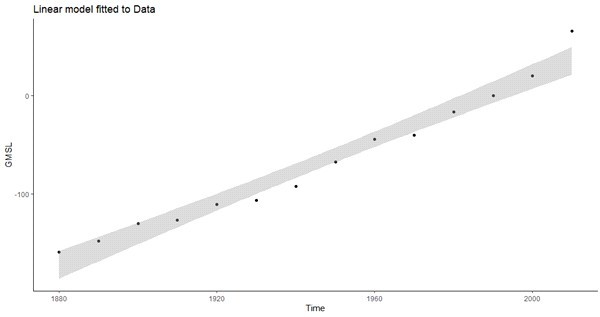
(Intercept) -3.175e+03 1.639e+02 -19.37 2.03e-10 \*\*\* Time 1.597e+00 8.426e-02 18.96 2.60e-10 \*\*\*

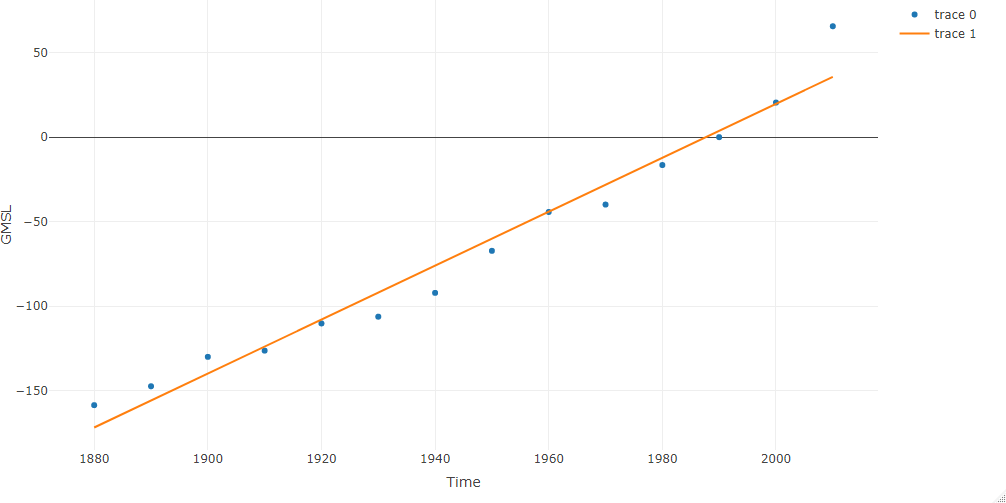
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.71 on 12 degrees of freedom Multiple R-squared: 0.9677, Adjusted R-squared: 0.965

F-statistic: 359.4 on 1 and 12 DF, p-value: 2.604e-10





## Prediction:

* new\_year<-data.frame(Time=seq(2010,2100,10))
* predict(spl,newdata=new\_year)

1 2 3 4 5 6 7 8 9 10

61.94048 99.77428 137.60809 175.44189 213.27570 251.10950 288.94331 326.77711 364.61091

402.44472

## CO2 level Vs Year:

* 1. **Spline Regression:**

summary(model)

Family: gaussian

Link function: identity Formula:

Total ~ s(Year) Parametric coefficients:

Estimate Std. Error t value Pr(&gt;|t|) (Intercept) 1491.46 11.85 125.9 &lt;2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Approximate significance of smooth terms:

edf Ref.df F p-value

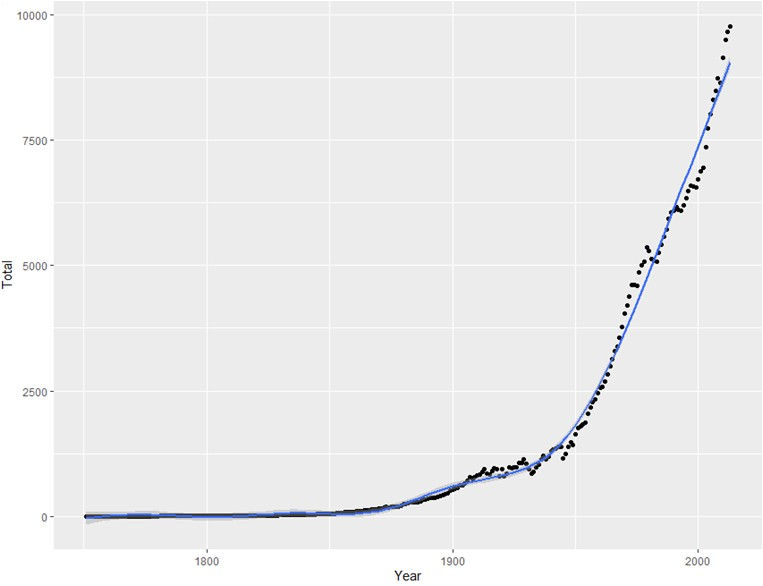
s(Year) 8.56 8.945 4517 &lt;2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 R-sq.(adj) = 0.994 Deviance explained = 99.4%

GCV = 38300 Scale est. = 36908 n = 263

ggplot(rio\_xlsx,aes(Year,Total))+geom\_point()+stat\_smooth(method=gam,formula= y~s(x))



## LOESS:

summary(gam(y~x,data=rio\_xlsx))

Family: gaussian

Link function: identity Formula:

y ~ x

Parametric coefficients:

Estimate Std. Error t value Pr(&gt;|t|) (Intercept) -36.301967 156.778237 -0.232 0.818

xYear 0.018318 0.080725 0.227 0.821

xGas 0.996948 0.004320 230.793 &lt;2e-16 \*\*\*

xLiquids 1.000776 0.001830 546.921 &lt;2e-16 \*\*\*

xSolids 1.000045 0.001514 660.684 &lt;2e-16 \*\*\*

xProduction 1.004212 0.009340 107.513 &lt;2e-16 \*\*\*

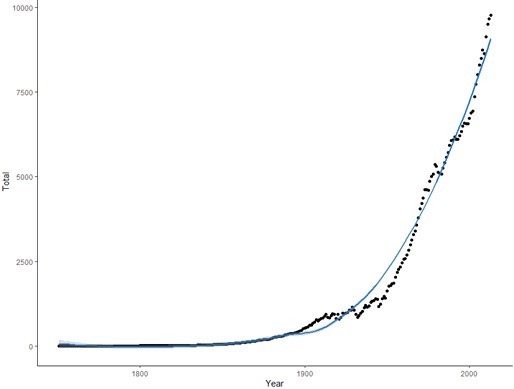
xFlaring 0.982561 0.020330 48.331 &lt;2e-16 \*\*\*

xCapita 0.971447 5.221717 0.186 0.853

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) = 0.934 Deviance explained = 93.7% GCV = 0.58108 Scale est. = 0.50845 n = 64



## Step Wise linear regression: 1)Forward:

summary(step2.model)

Call:

lm(formula = Total ~ Year, data = rio\_xlsx)

Residuals:

Min 1Q Median 3Q Max

-1900.5 -1320.8 -295.6 1060.5 5021.8

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -45381.72 2242.30 -20.24 <2e-16 \*\*\*

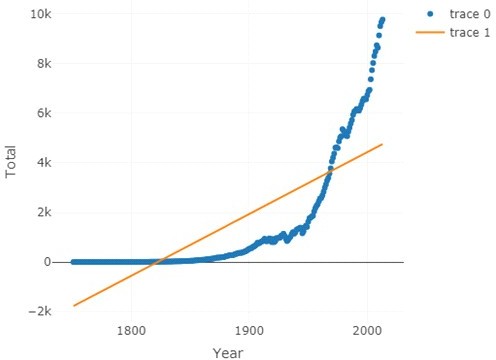
Year 24.91 1.19 20.92 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1466 on 261 degrees of freedom Multiple R-squared: 0.6264, Adjusted R-squared: 0.625

F-statistic: 437.7 on 1 and 261 DF, p-value: < 2.2e-16



## Backward:

summary(step3.model)

Call:

lm(formula = Total ~ Year, data = rio\_xlsx)

Residuals:

Min 1Q Median 3Q Max

-1900.5 -1320.8 -295.6 1060.5 5021.8

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) -45381.72 2242.30 -20.24 <2e-16 \*\*\*

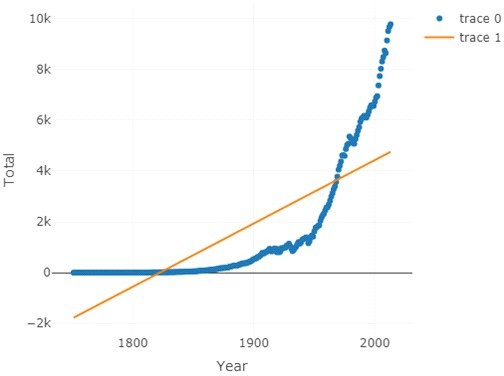
Year 24.91 1.19 20.92 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1466 on 261 degrees of freedom Multiple R-squared: 0.6264, Adjusted R-squared: 0.625

F-statistic: 437.7 on 1 and 261 DF, p-value: < 2.2e-16



## Step wise selection:

summary(step1.model)

Call:

lm(formula = Total ~ Year, data = rio\_xlsx)

Residuals:

Min 1Q Median 3Q Max

-1900.5 -1320.8 -295.6 1060.5 5021.8

Coefficients:

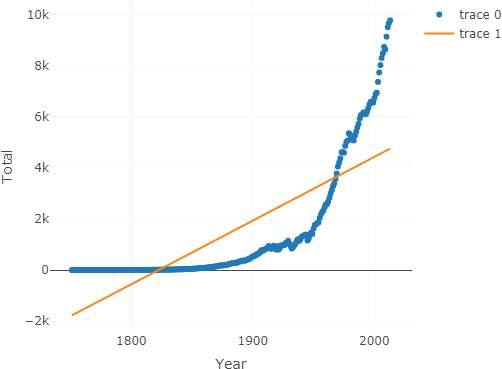
Estimate Std. Error t value Pr(>|t|) (Intercept) -45381.72 2242.30 -20.24 <2e-16 \*\*\*

Year 24.91 1.19 20.92 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1466 on 261 degrees of freedom Multiple R-squared: 0.6264, Adjusted R-squared: 0.625 F-statistic: 437.7 on 1 and 261 DF, p-value: < 2.2e-16



## Power 4 and Power 5:

Power 4:

summary(m4)

Call:

lm(formula = Total ~ poly(Year, 4, raw = TRUE), data = rio\_xlsx)

Residuals:

Min 1Q Median 3Q Max

-688.37 -36.99 -0.37 54.63 861.90

Coefficients:

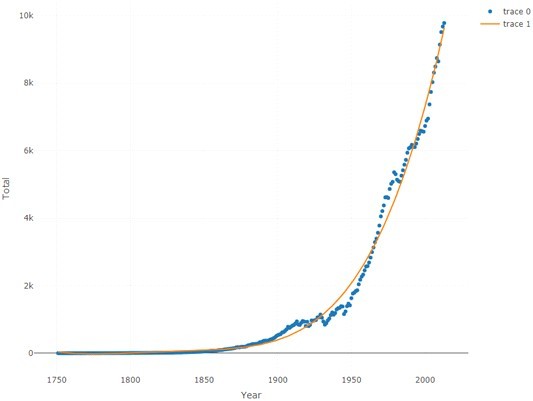
Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.569e+07 7.736e+06 9.785 <2e-16 \*\*\* poly(Year, 4, raw = TRUE)1 -1.664e+05 1.647e+04 -10.103 <2e-16 \*\*\* poly(Year, 4, raw = TRUE)2 1.372e+02 1.314e+01 10.438 <2e-16 \*\*\* poly(Year, 4, raw = TRUE)3 -5.027e-02 4.659e-03 -10.790 <2e-16 \*\*\* poly(Year, 4, raw = TRUE)4 6.906e-06 6.189e-07 11.159 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 228.6 on 258 degrees of freedom Multiple R-squared: 0.991, Adjusted R-squared: 0.9909 F-statistic: 7117 on 4 and 258 DF, p-value: < 2.2e-16



Power 5:

summary(m5)

Call:

lm(formula = Total ~ poly(Year, 5, raw = TRUE), data = rio\_xlsx)

Residuals:

Min 1Q Median 3Q Max

-688.37 -36.99 -0.37 54.63 861.90

Coefficients: (1 not defined because of singularities) Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.569e+07 7.736e+06 9.785 <2e-16 \*\*\*

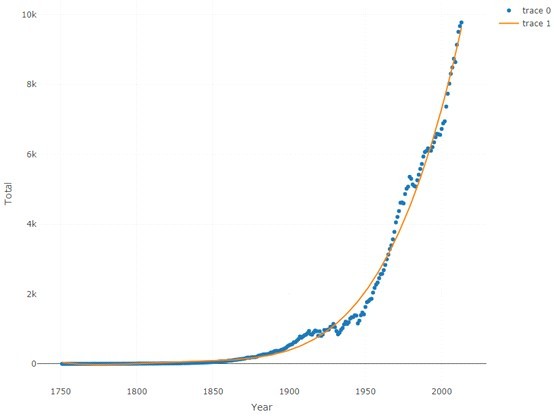
poly(Year, 5, raw = TRUE)1 -1.664e+05 1.647e+04 -10.103 <2e-16 \*\*\* poly(Year, 5, raw = TRUE)2 1.372e+02 1.314e+01 10.438 <2e-16 \*\*\* poly(Year, 5, raw = TRUE)3 -5.027e-02 4.659e-03 -10.790 <2e-16 \*\*\* poly(Year, 5, raw = TRUE)4 6.906e-06 6.189e-07 11.159 <2e-16 \*\*\*

poly(Year, 5, raw = TRUE)5 NA NA NA NA

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 228.6 on 258 degrees of freedom

Multiple R-squared: 0.941, Adjusted R-squared: 0.9509 F-statistic: 7117 on 4 and 258 DF, p-value: < 2.2e-16

## Cubic and Square:

Cubic:

summary(m3)

Call:

lm(formula = Total ~ poly(Year, 3, raw = T), data = rio\_xlsx)

Residuals:

Min 1Q Median 3Q Max

-751.53 -182.90 13.56 218.56 687.21

Coefficients:

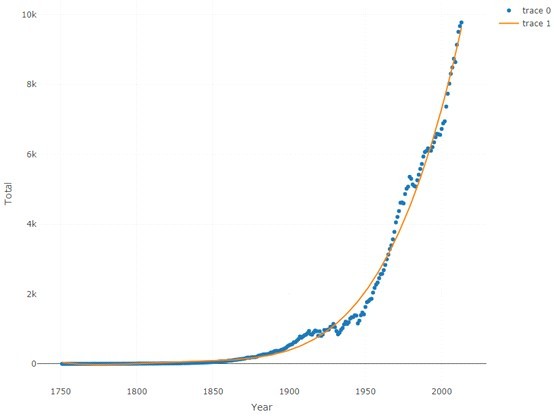
Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.058e+07 3.314e+05 -31.93 <2e-16 \*\*\* poly(Year, 3, raw = T)1 1.733e+04 5.292e+02 32.75 <2e-16 \*\*\* poly(Year, 3, raw = T)2 -9.458e+00 2.814e-01 -33.61 <2e-16 \*\*\* poly(Year, 3, raw = T)3 1.720e-03 4.984e-05 34.51 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 277.8 on 259 degrees of freedom Multiple R-squared: 0.9867, Adjusted R-squared: 0.9865 F-statistic: 6396 on 3 and 259 DF, p-value: < 2.2e-16



Square:

summary(m2)

Call:

lm(formula = Total ~ poly(Year, 2, raw = T), data = rio\_xlsx)

Residuals:

Min 1Q Median 3Q Max

-1445.7 -481.3 143.2 456.8 2133.9

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 8.507e+05 2.777e+04 30.63 <2e-16 \*\*\*

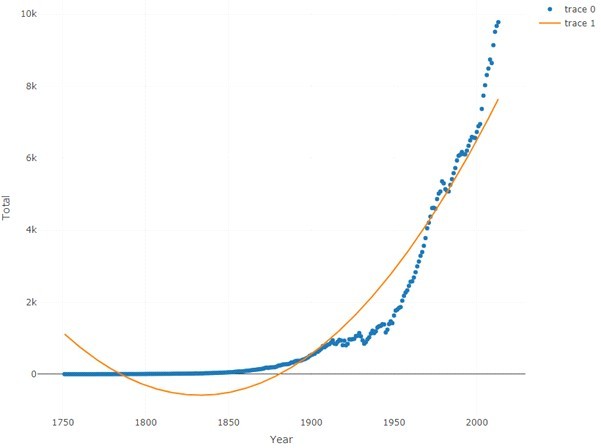
poly(Year, 2, raw = T)1 -9.289e+02 2.954e+01 -31.44 <2e-16 \*\*\* poly(Year, 2, raw = T)2 2.534e-01 7.848e-03 32.29 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 656.1 on 260 degrees of freedom

Multiple R-squared: 0.9254, Adjusted R-squared: 0.9249 F-statistic: 1613 on 2 and 260 DF, p-value: < 2.2e-16



## Exponential:

summary(m21) Call:

lm(formula = Total ~ poly(Year, 2.732, raw = T), data = rio\_xlsx)

Residuals:

Min 1Q Median 3Q Max

-1445.7 -481.3 143.2 456.8 2133.9

Coefficients:

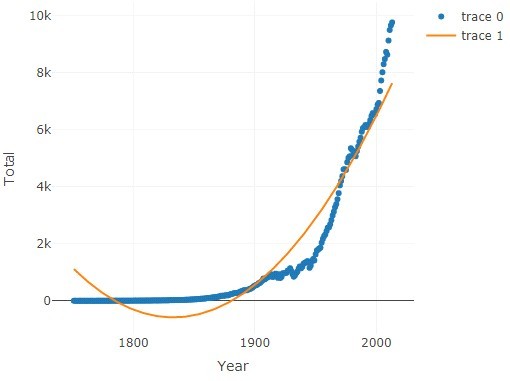
Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.507e+05 2.777e+04 30.63 <2e-16 \*\*\* poly(Year, 2.732, raw = T)1 -9.289e+02 2.954e+01 -31.44 <2e-16 \*\*\* poly(Year, 2.732, raw = T)2 2.534e-01 7.848e-03 32.29 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Residual standard error: 656.1 on 260 degrees of freedom

Multiple R-squared: 0.9254, Adjusted R-squared: 0.9249 F-statistic: 1613 on 2 and 260 DF, p-value: < 2.2e-16



## Prediction:

