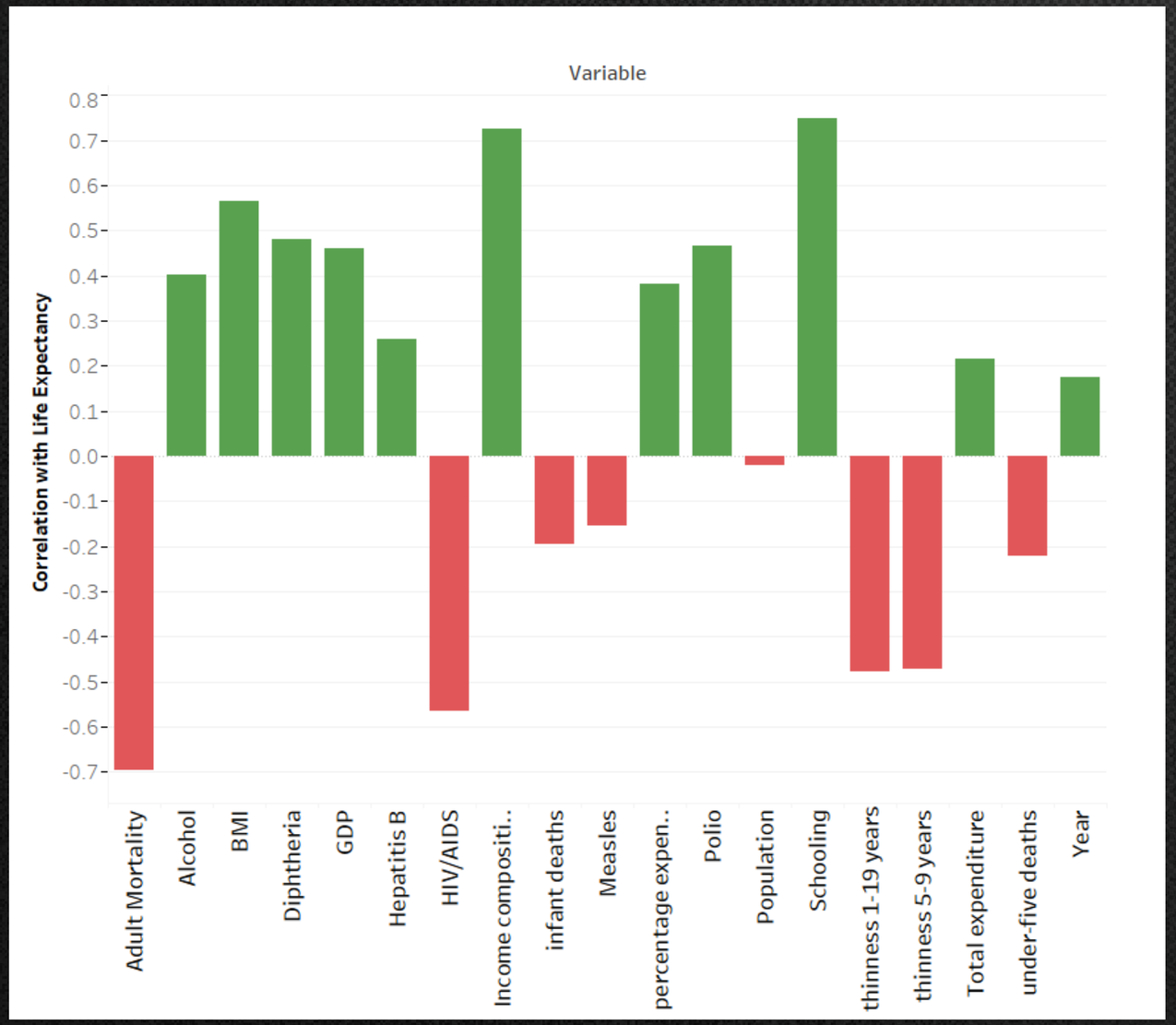
No Time to Die

BY STRAW HATS

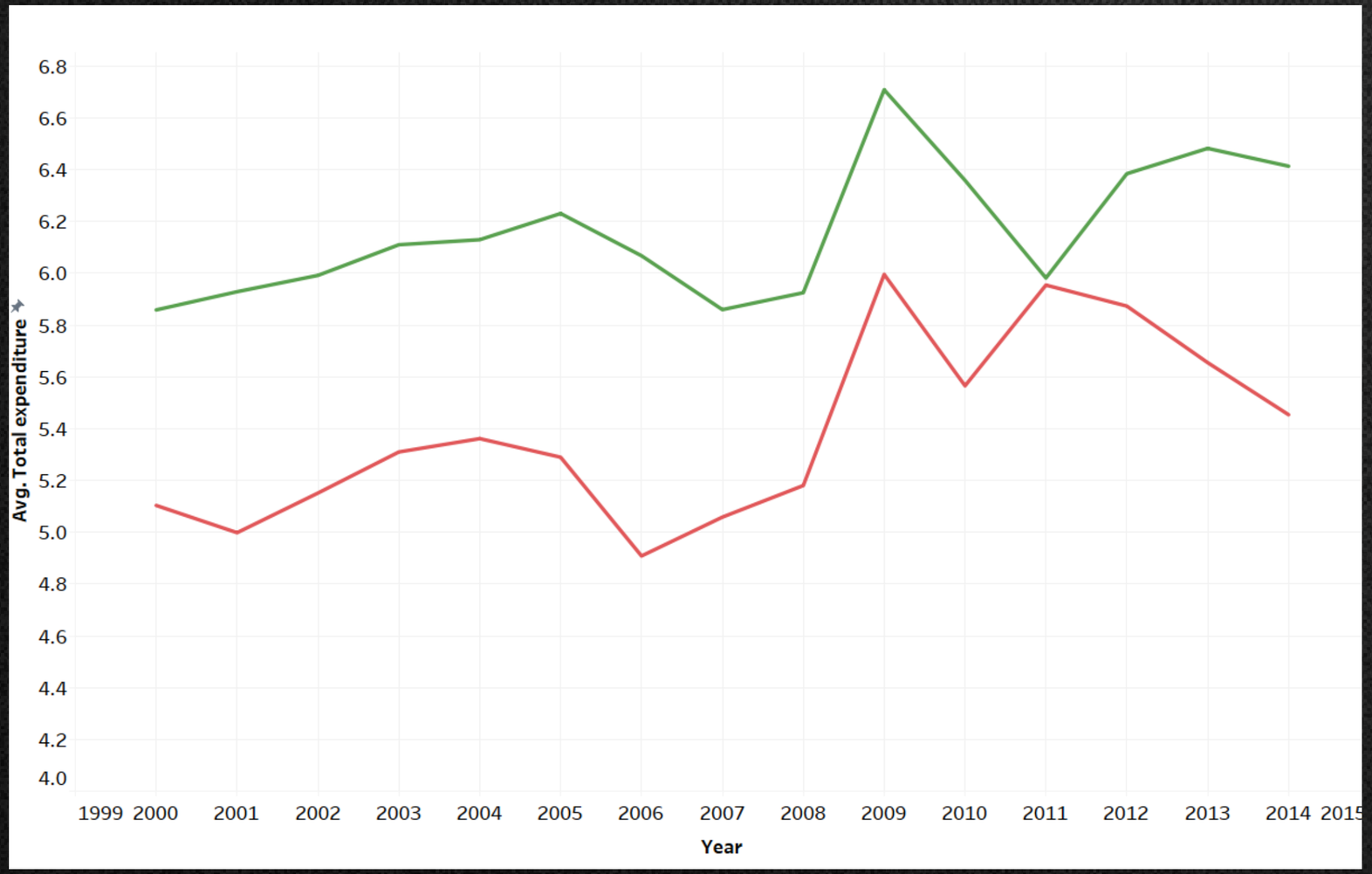
Life Expectency Prediction and Statistical Analysis



Factors

What really affects Life Expectancy?

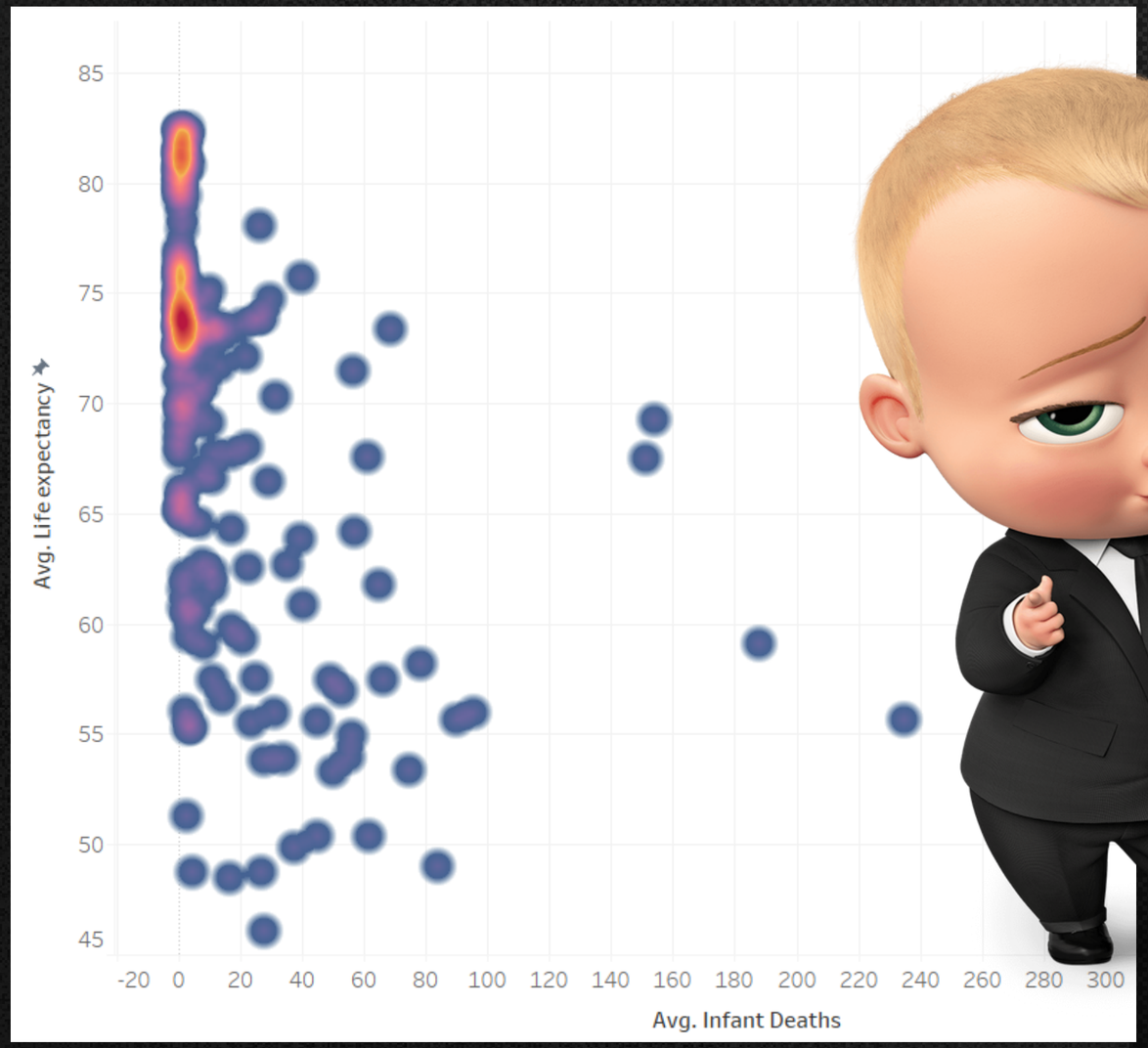
Should a country invest more?

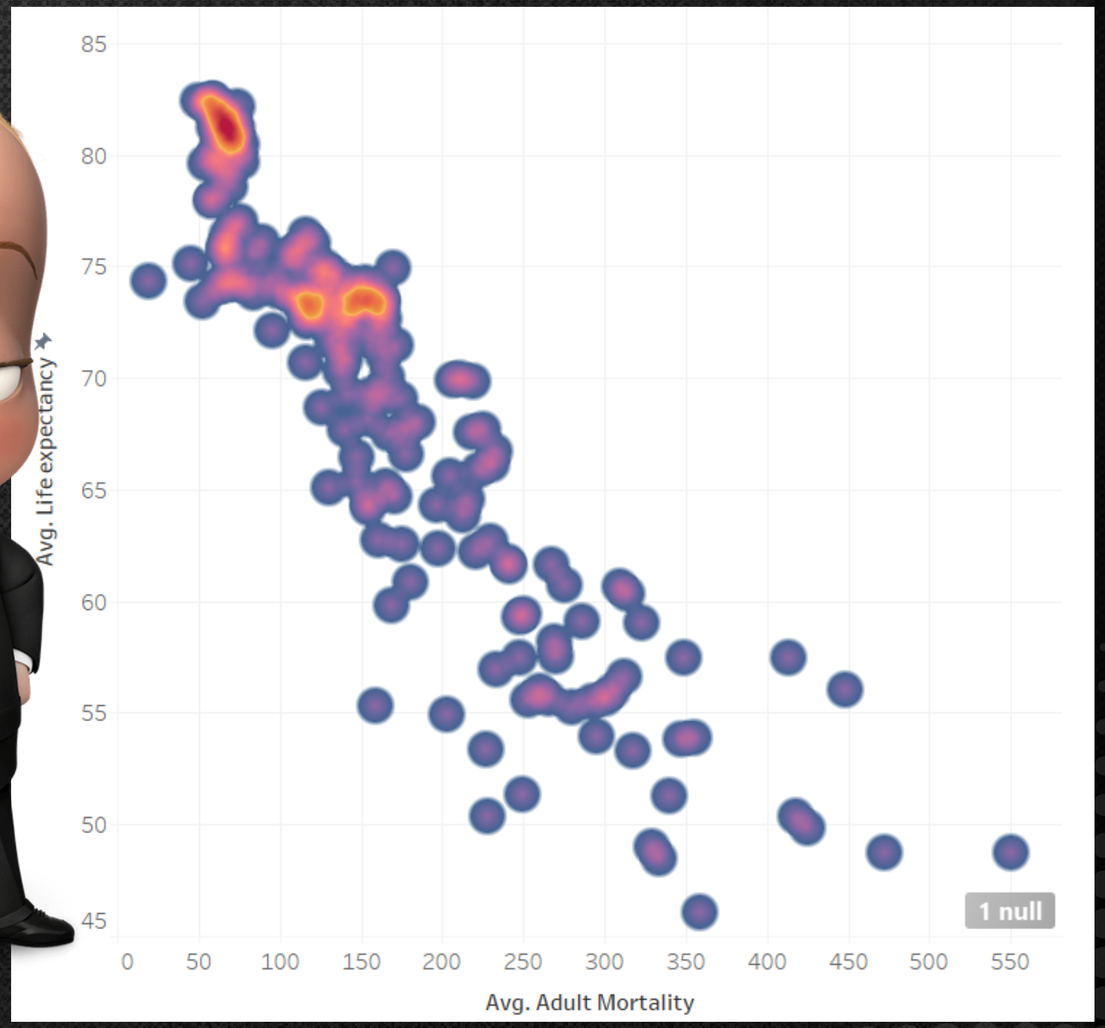


Healthcare Expenditure

Infants

Adults

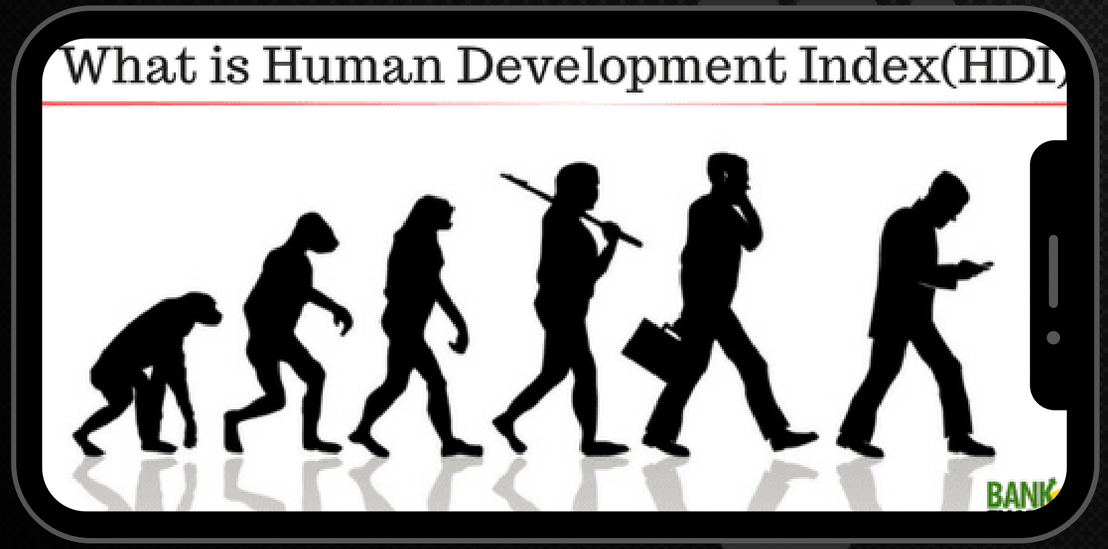


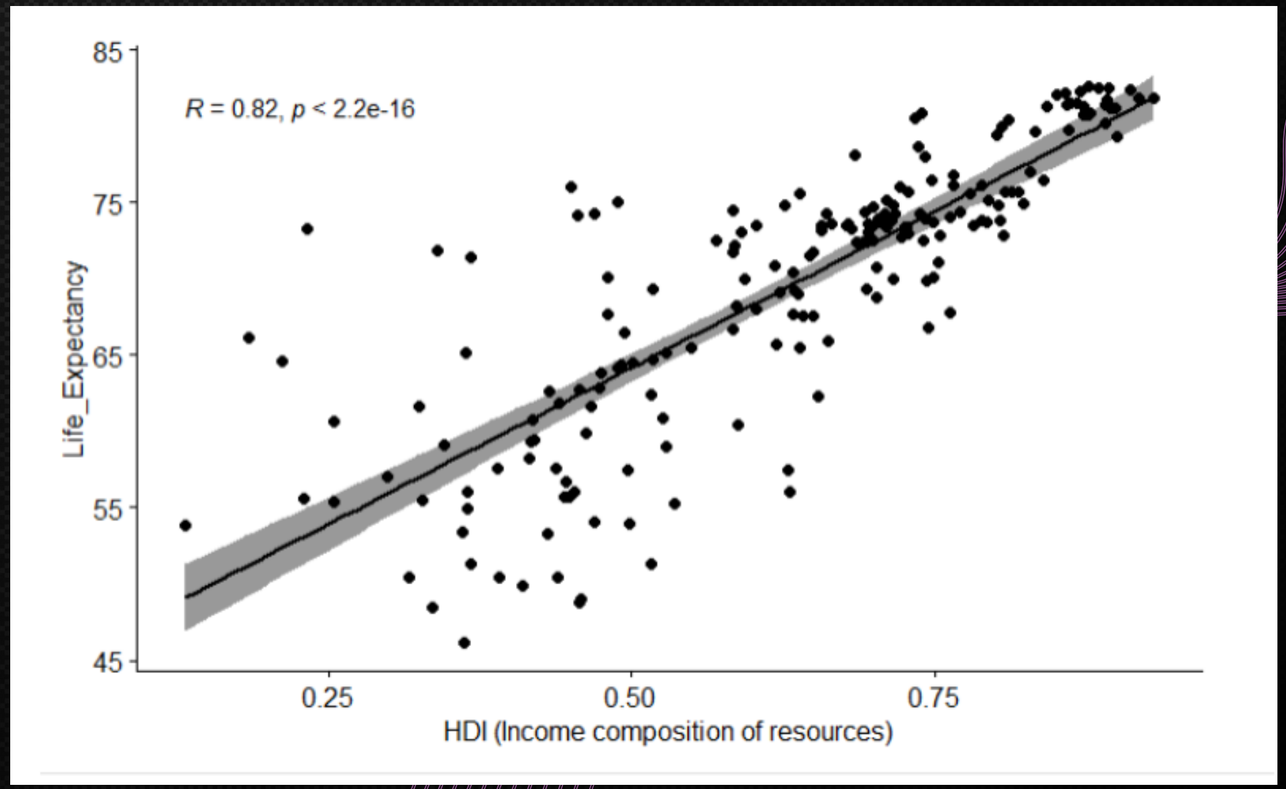


```
        Pearson's product-moment correlation

data:  data1$Average_HDI and data1$Average_life
t = 19.515, df = 191, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7627543 0.8583782
sample estimates:
      cor
0.8160772
```

Why is HDI?



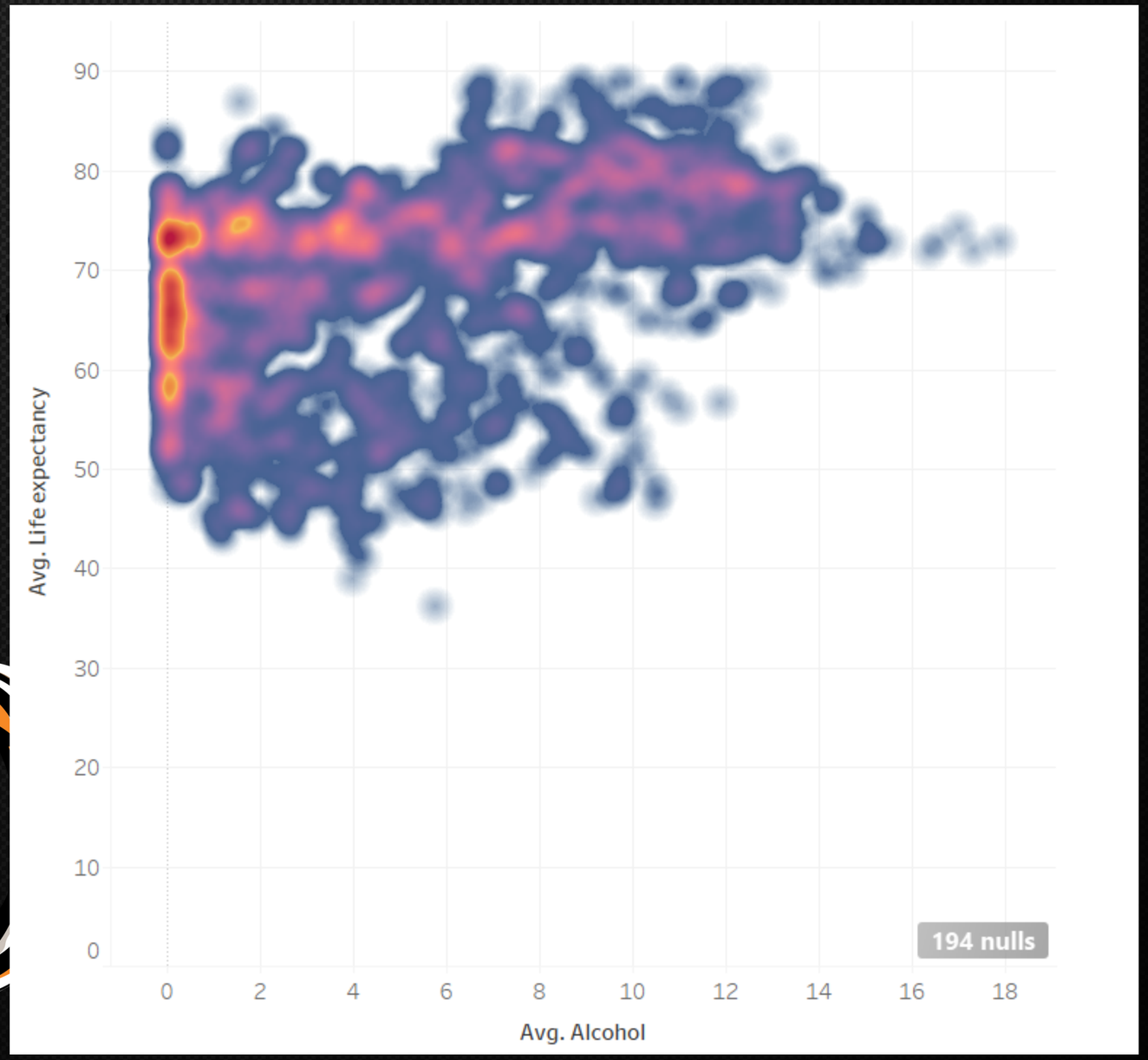


Anova on Schooling

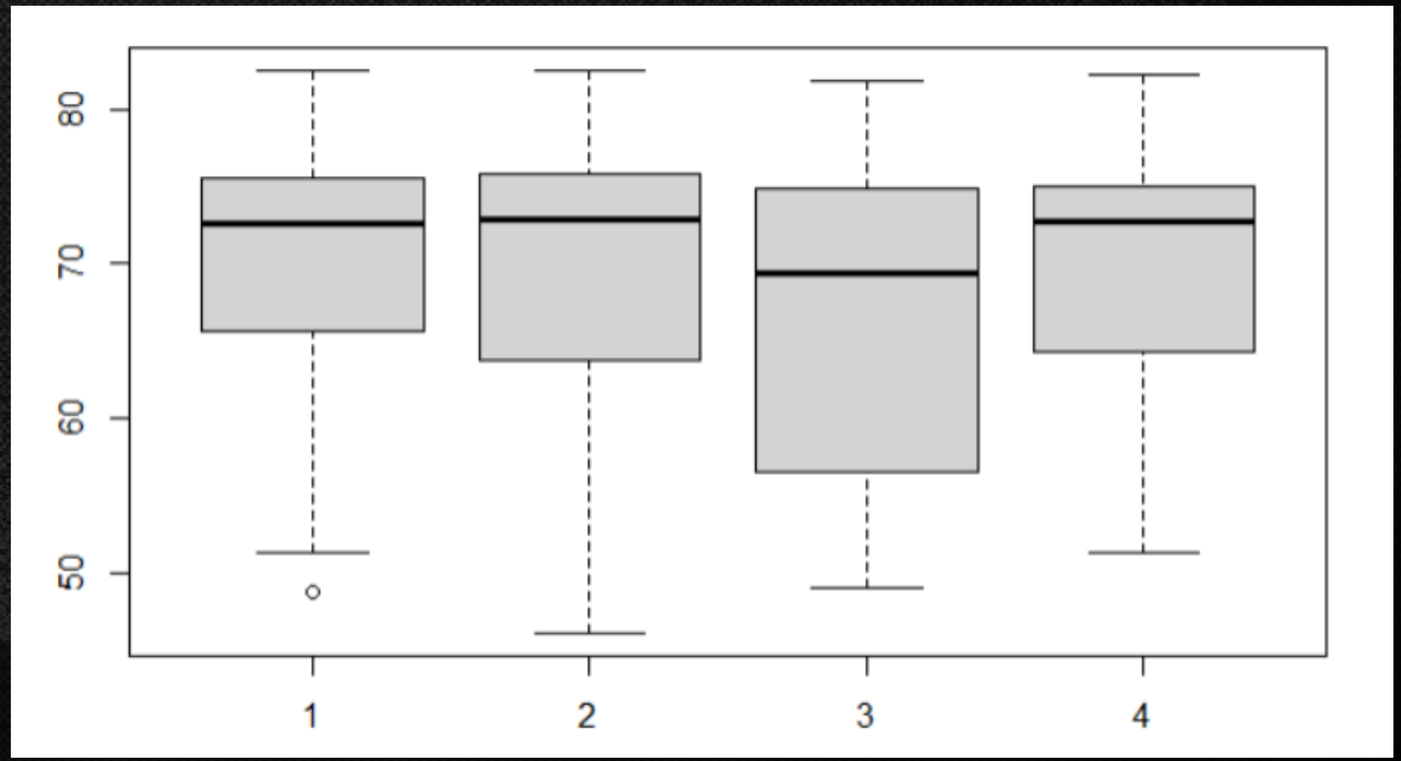
```
             Df Sum Sq Mean Sq F value Pr(>F)
Education     2   8627    4313   114.4 <2e-16 ***
Residuals   190   7166      38
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

As the P value is less than 0.05. That means, there is a significant difference and it does create an impact.

To Drink or not to Drink?



As the population increases, we see that the median life expectancy decreases, as seen in the third boxplot. At first it seems counterintuitive as to why the expectancy increases in the fourth boxplot. But this is because the categories only refer to the population and not population density.



Immunity!

```
             Df Sum Sq Mean Sq F value   Pr(>F)
Polio         1   4855    4855   98.59  < 2e-16 ***
Diphtheria    1   1581    1581   32.10 5.35e-08 ***
Residuals   190   9357      49
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
             Df Sum Sq Mean Sq F value   Pr(>F)
Hepatitis     1   2865  2865.0   42.33 6.61e-10 ***
Residuals   191  12928    67.7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Aproach

aka Our Game Plan

Observe

UNDERSTAND WHY

Predict

USE ML TO PREDICT

Examine

WHAT IS THE PROBLEM?

Prescribe

HOW CAN WE SOLVE IT?

Our Track

1

Cleaning

Data cleaning to handle missing values and remove duplicates.

2

Preprocessing

Dimentionality Reduction, Scaling

3

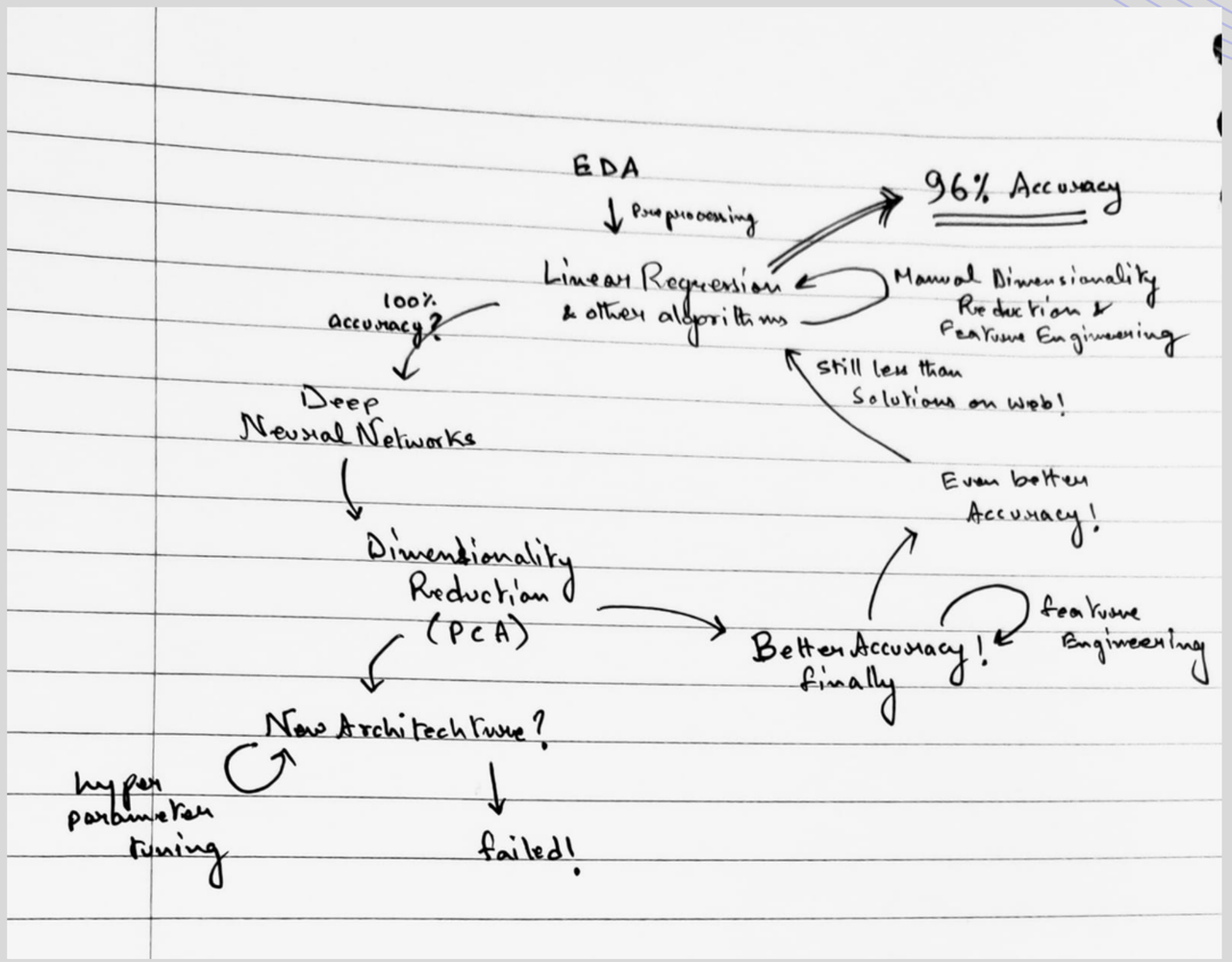
Modeling

Various models using different algorithms.

4

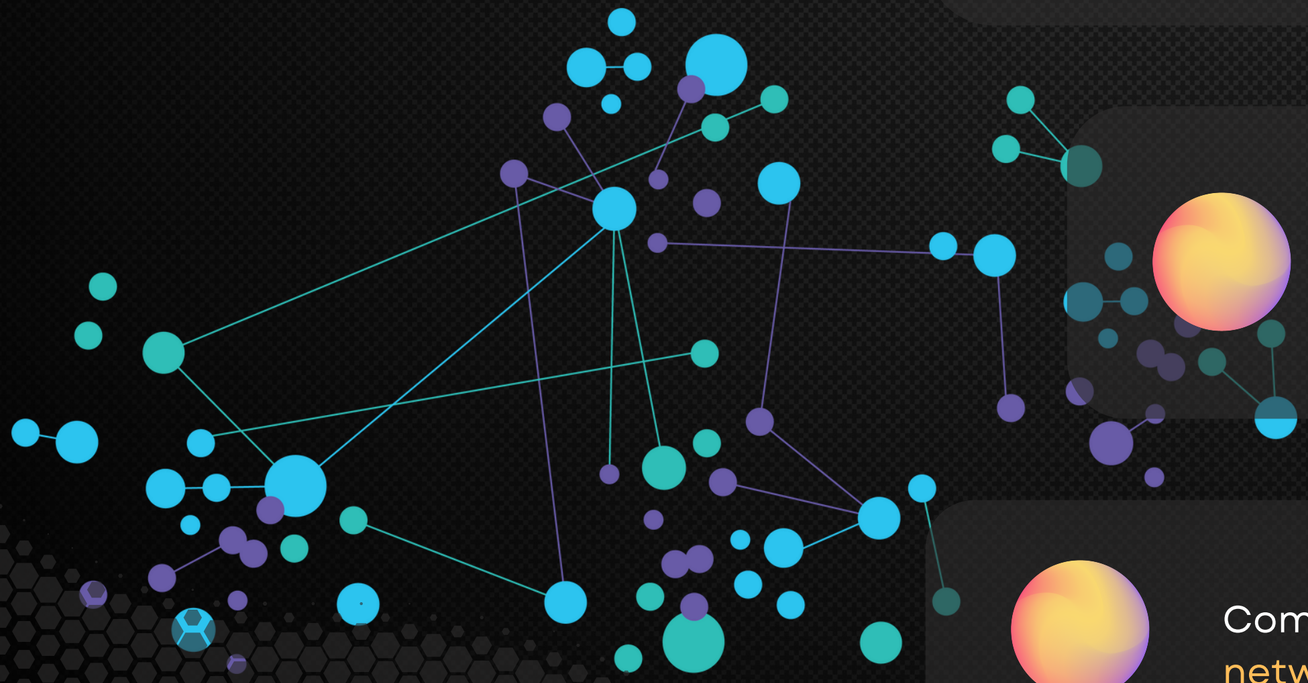
Feature Engineering

Getting the most out of the data


EDA
Preprocessing
96% Accuracy
Linear Regression & other algorithms
Manual Dimensionality Reduction & Feature Engineering
100% accuracy?
Still less than Solutions on Web!
Deep Neural Networks
Even better Accuracy!
Dimensionality Reduction (PCA)
Better Accuracy! finally
feature Engineering
New Architecture?
hyper parameter tuning
failed!

Prediction Models

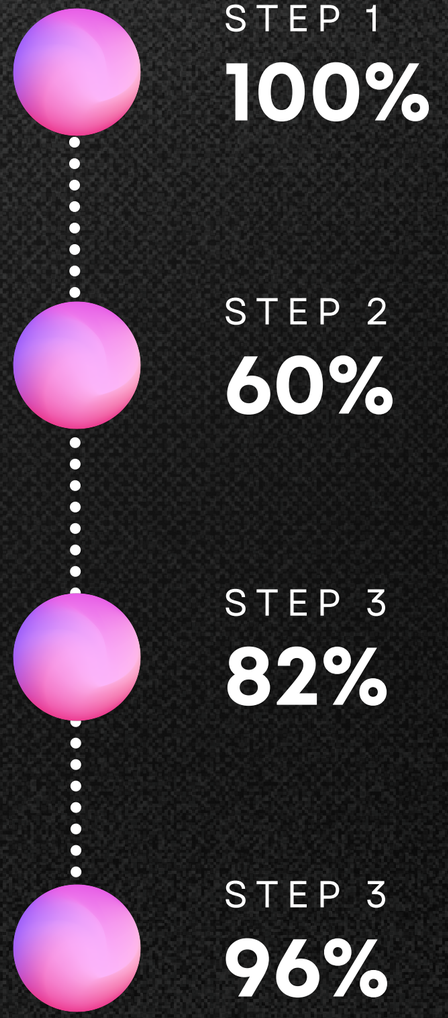
Simplicity over Complexity



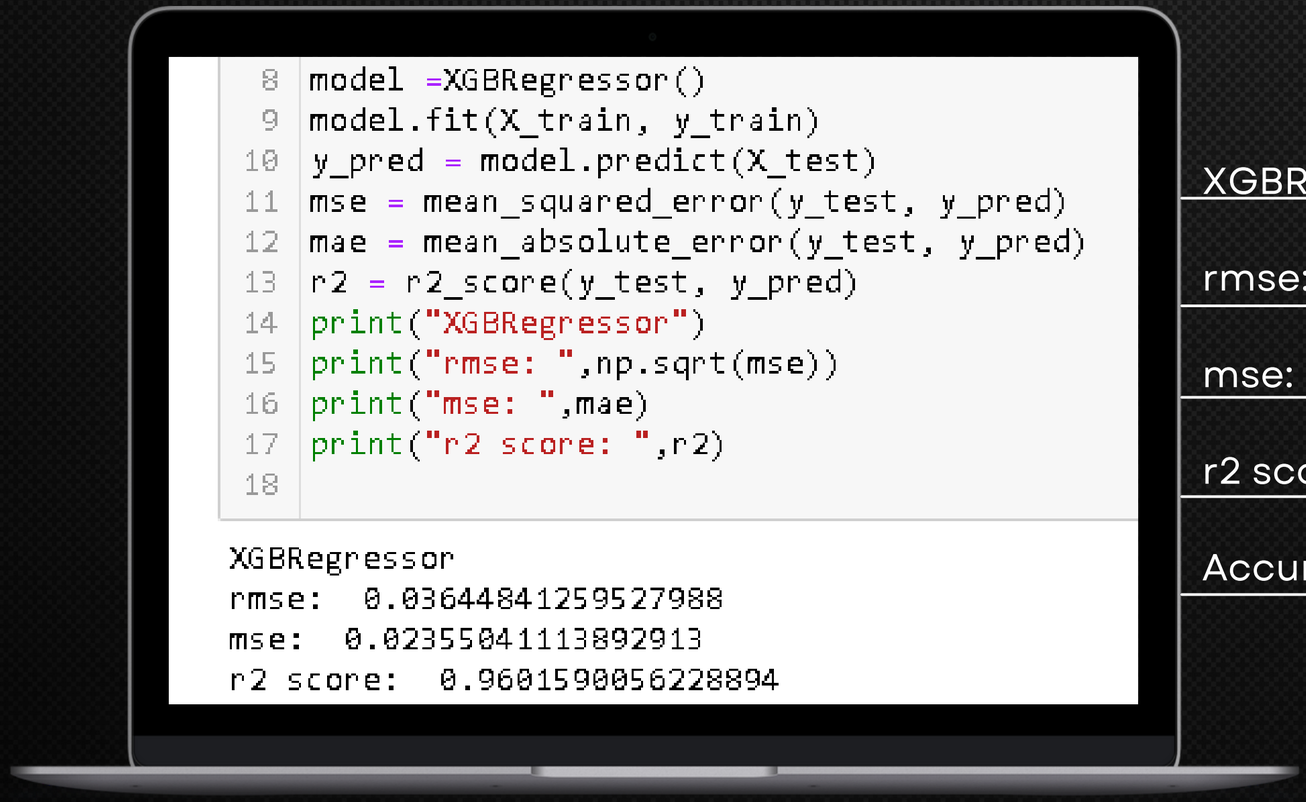
- Comparitive study of different Algorithms - Suspected Overfitting.
- Transition to Deep Learning and experience with PCA.
- Coming back to roots with natural neural networks.

Accuracy Bumps

accuracy will increase and decrease, but what stays constant is the will to get a better solution



XGB Regressor



```
model =XGBRegressor()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print("XGBRegressor")
print("rmse: ",np.sqrt(mse))
print("mse: ",mae)
print("r2 score: ",r2)

```

```
XGBRegressor
rmse:  0.03644841259527988
mse:  0.02355041113892913
r2 score:  0.9601590056228894
```

XGBRegressor

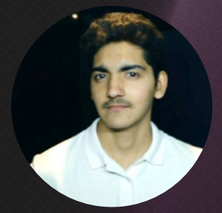
rmse: 0.03644

mse: 0.023550

r2 score: 0.96015

Accuracy: 96.015%

The Team

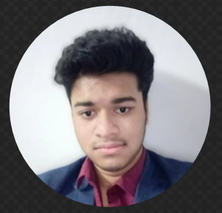


Ansh Sancheti

DATA ANALYST

Akshay Swami

ML ENGINEER



Nirmalya Sarkar

ML ENGINEER

Sandeep Jabez

WEB APP DEVELOPER

Thank You

STRAW HATS