

# DH 302 MIDSEM PROJECT

CHIRAG, 18B090003

AMEYA, 19B090002



# CARDIOVASCULAR DISEASE

- Cardi(o) – pertaining to the heart
- Vascular – pertaining to arteries, veins, etc

# Why study cardiovascular disease?

- Cardiovascular disease is one of the most severe chronic diseases and is the leading cause of death and disability in the world.
- Chronic diseases are defined as conditions that last for 1 year or more and require ongoing medical attention or limit activities of daily living or both – other types are diabetes, alzheimers, cancer, obesity, etc.
- 17.9 million people die due to cardiovascular diseases each year.
- Heart disease costs the United States about \$219 billion each year. (This includes the cost of health care services, medicines, and lost productivity due to death)

- TYPES

**Broad types:**

Cerebrovascular disease(stroke)

Coronary artery disease

Congestive heart failure

Ischemic heart disease, Other diseases of the heart

MAY BE

Genetically predisposed, Environmentally acquired

# CORONARY ARTERY DISEASE

- Leading cause of heart attacks – **myocardial infarctions**.
- Occurs when plaque builds up in the arteries, narrowing it and limiting blood flow to the heart. Without blood, tissue loses oxygen and dies.
- Causes:
  - High blood pressure
  - Eating food high in cholesterol
  - Smoking
  - Lack of regular exercise
  - Diabetes

# CONGESTIVE HEART FAILURE

Heart is unable to pump adequate amount of blood throughout the body.

- Causes:
  - Obesity , high cholesterol level, high blood pressure
  - Unhealthy (high salt) diet
  - Stress
  - Certain medications like the syrup of Ipecac which is myotoxic used by people with eating disorders
- Treatment: Adding a pacemaker (generates electrical impulses delivered by electrodes to cause the heart muscle chambers to contract and therefore pump blood

# CEREBROVASCULAR DISEASE/STROKE

- Damage to the brain from interruption of its blood supply, which causes brain cells to die.  
**Ischemic** stroke and **Haemorrhagic** stroke
- Symptoms of stroke include trouble walking, speaking and understanding, as well as paralysis or numbness of the face, arm or leg.

## Causes

- Fatty deposits clog the brain's blood vessels because of **high cholesterol**
- Rupture in a blood vessel / blood clot / embolism

# TREATMENT FOR HEART DISEASE:

- **Lifestyle changes**

A factoid: **Bankers** have a 10% higher chance of a heart-related event

- **Medication**

Blood thinners,  $\beta$  – blockers, and so on

- **Implants**

Stents, Pacemakers, Artificial valves and so on

- **Surgery**

ranges from angioplasty to open heart surgery



# RISKS

- Family history
- Race / Ethnicity
  - South Asians are more likely to develop coronary heart disease than white Europeans.
  - African or African Caribbean people are at higher risk of developing high blood pressure and having a stroke than other ethnic groups.
  - In the US, highest risk for obesity and death from heart diseases is seen in **black** people, followed by non-Hispanic **whites**.
  - Men are at greater risk for heart disease than women. According to the CDC, 70-89% of all cardiac events in the United States occur in men.

# WELCOME TO THE REAL WORLD

Dataset: <https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi-d1f3d>

Name: `"U.S._Chronic_Disease_Indicators__CDI_"`

Dimensions: **956638 \* 34**

(show excel file)

# About the data set

- In this crude data set, unorganised “**facts**” or “**data**” have been presented about **all** health related records. We will try to extract **information** about four conditions:
  - Coronary artery disease
  - Congestive heart failure
  - Cerebro-vascular disease (Stroke)
  - Other Diseases of the heart

# Questions

1. How common is cholesterol screening among adults (18+) in these states?  
And what is the **high cholesterol** prevalence in these states?
2. What is the prevalence of **high blood pressure** (hypertension) in the adult population in USA?
3. What is the **hospitalization rate** for stroke and acute myocardial infarction?

Can these questions be answered just by looking at the data set?  
No. That's where the computer comes in.

# Elephant in the room

**Race-wise and Gender-wise mortality rates due to the above 4 mentioned conditions as a measure out of 100,000 population:**

- Which race has a higher mortality rate due to the above 4 diseases?
- What is the SMR( Standardized mortality ratio) in the case of each race, when compared to the overall population?
- How does the mortality look like when we compare men and women?

# To address the elephant

- We will consider the following races.

i. White

ii. Black

iii. Asian

iv. Latin/Hispanic

v. Asia/Pacific

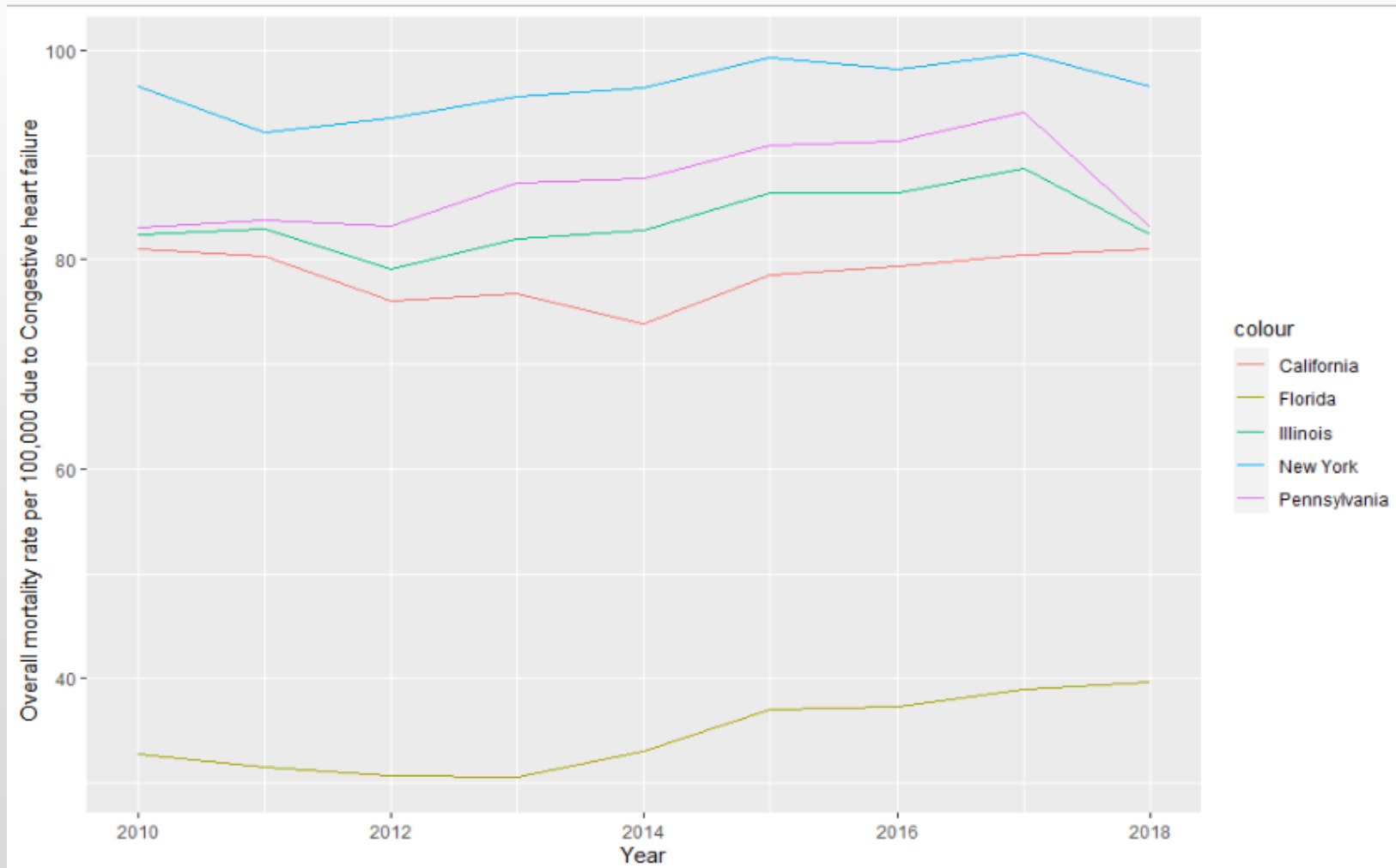
In the states of

California, New York  
And USA as a whole

# A few definitions

- **Age-adjusted mortality rate:** The difference in the ages of between any two population is normalized. In short, we assume equal age distribution in both the prospective populations.
- **Standardized Mortality Ratio:** This likelihood is expressed in terms of a factor  $n$  which is calculated using a standard formula ( using indirect age adjustment)
- **Prevalence:** It is a fraction that indicates the percentage of the population at risk living with a certain condition or ability. It can also be adjusted for age.

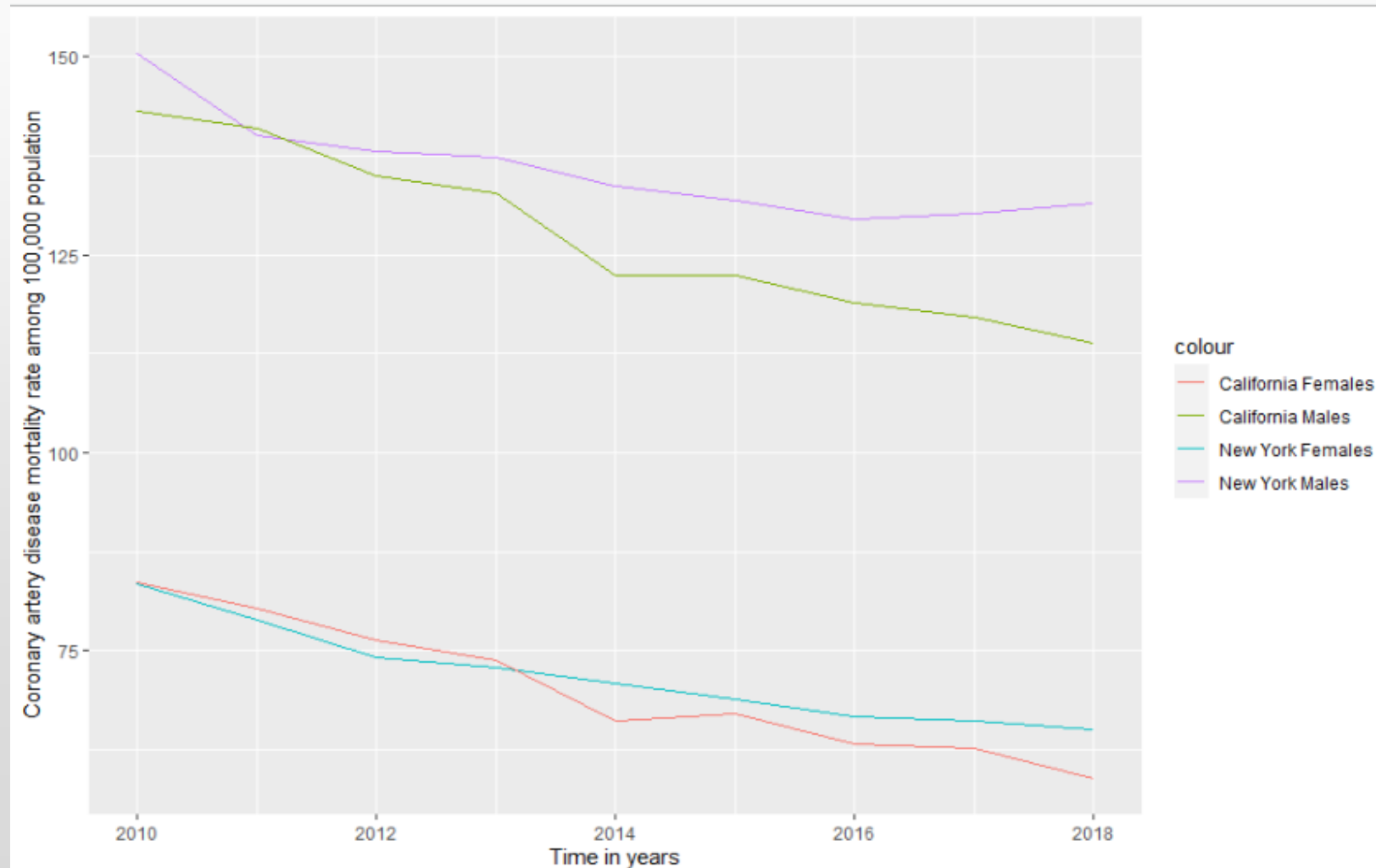
# Congestive heart failure - Mortality in 5 states (*Banker effect of New York??*)





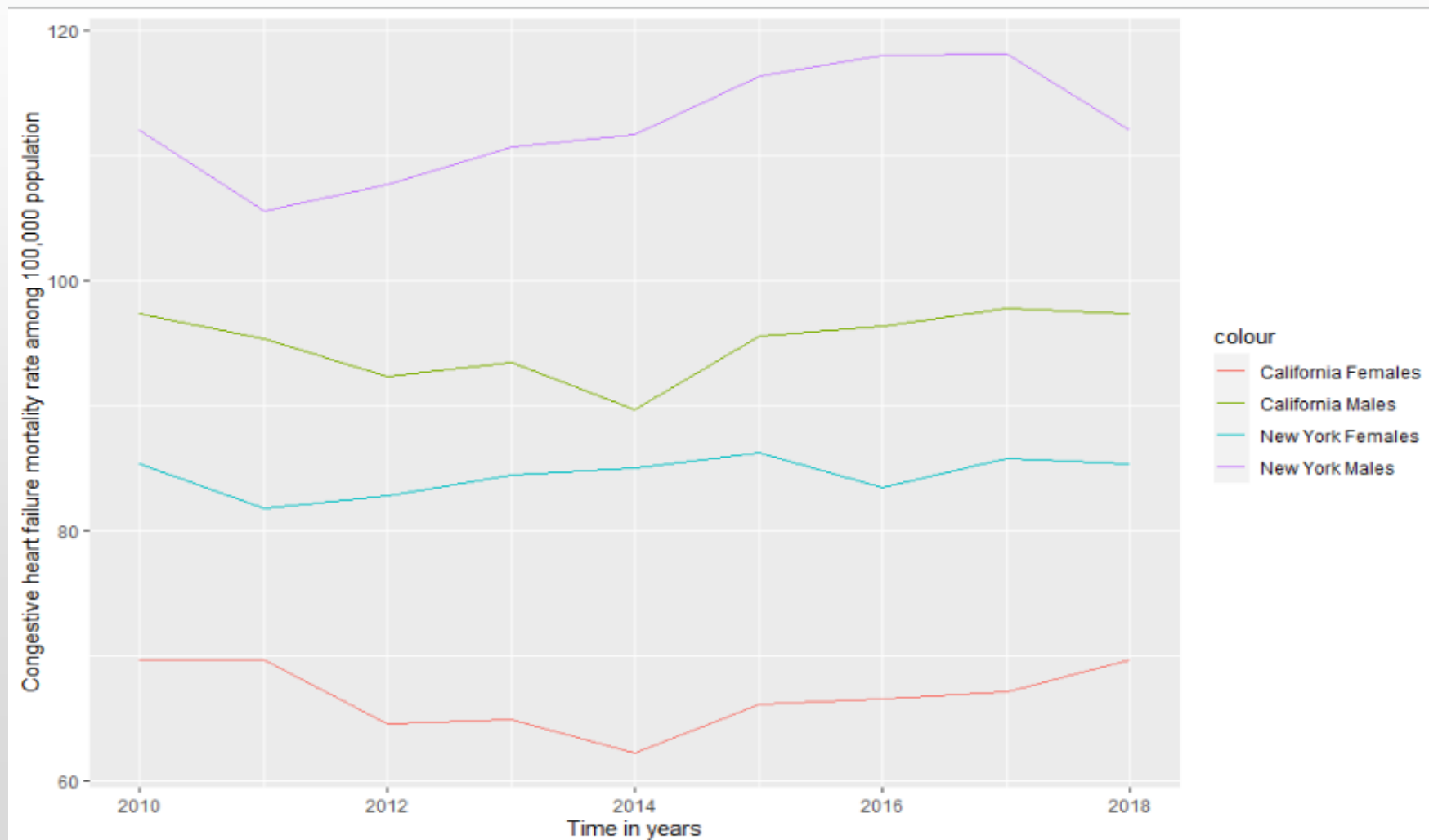
# MALE and FEMALE counterparts in two states

## Coronary artery disease



# MALE and FEMALE counterparts in two states

## Congestive heart failure



# Why do men have a higher risk of dying?

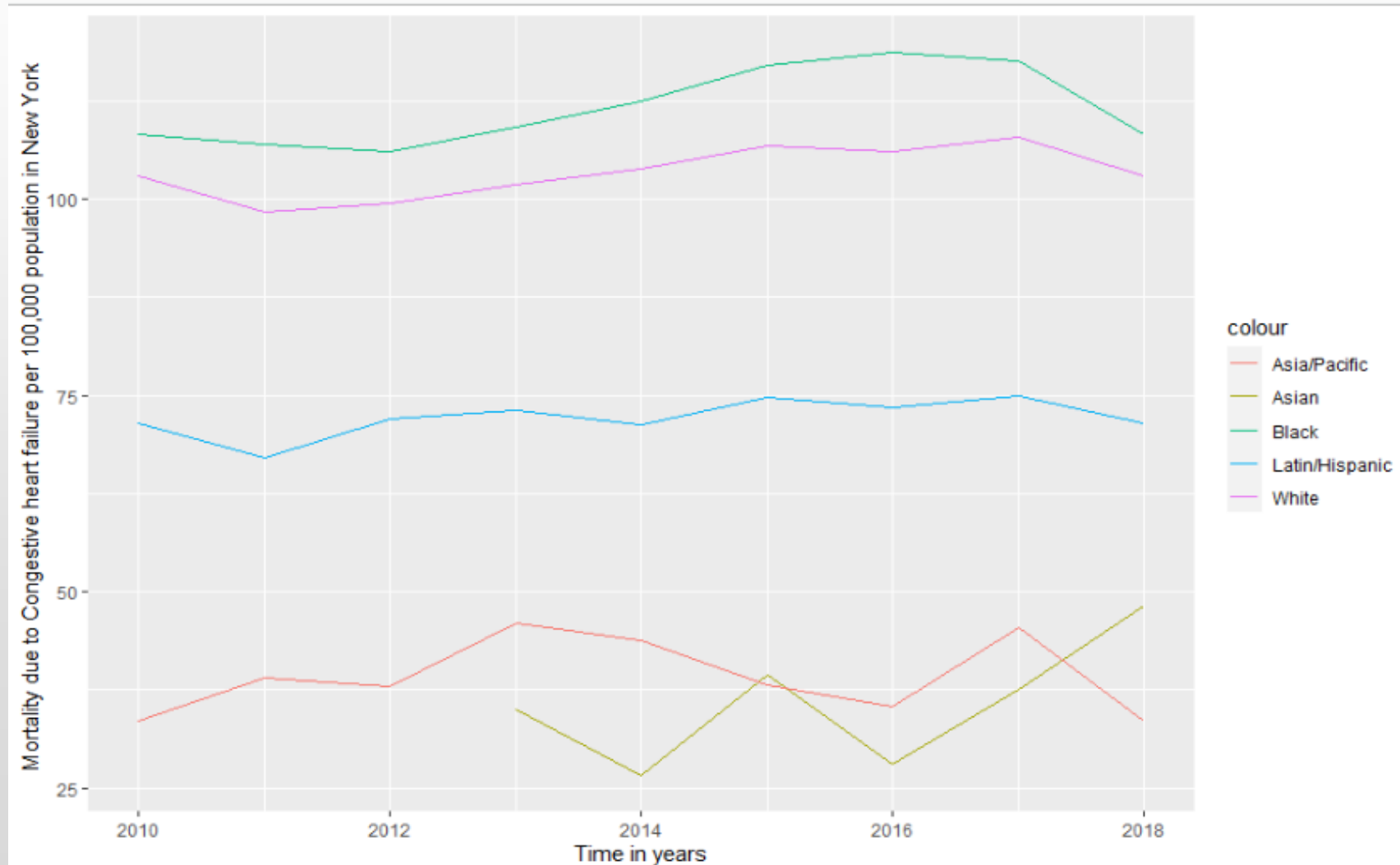
Is it because they live dangerously?

Maybe.

Probably not.

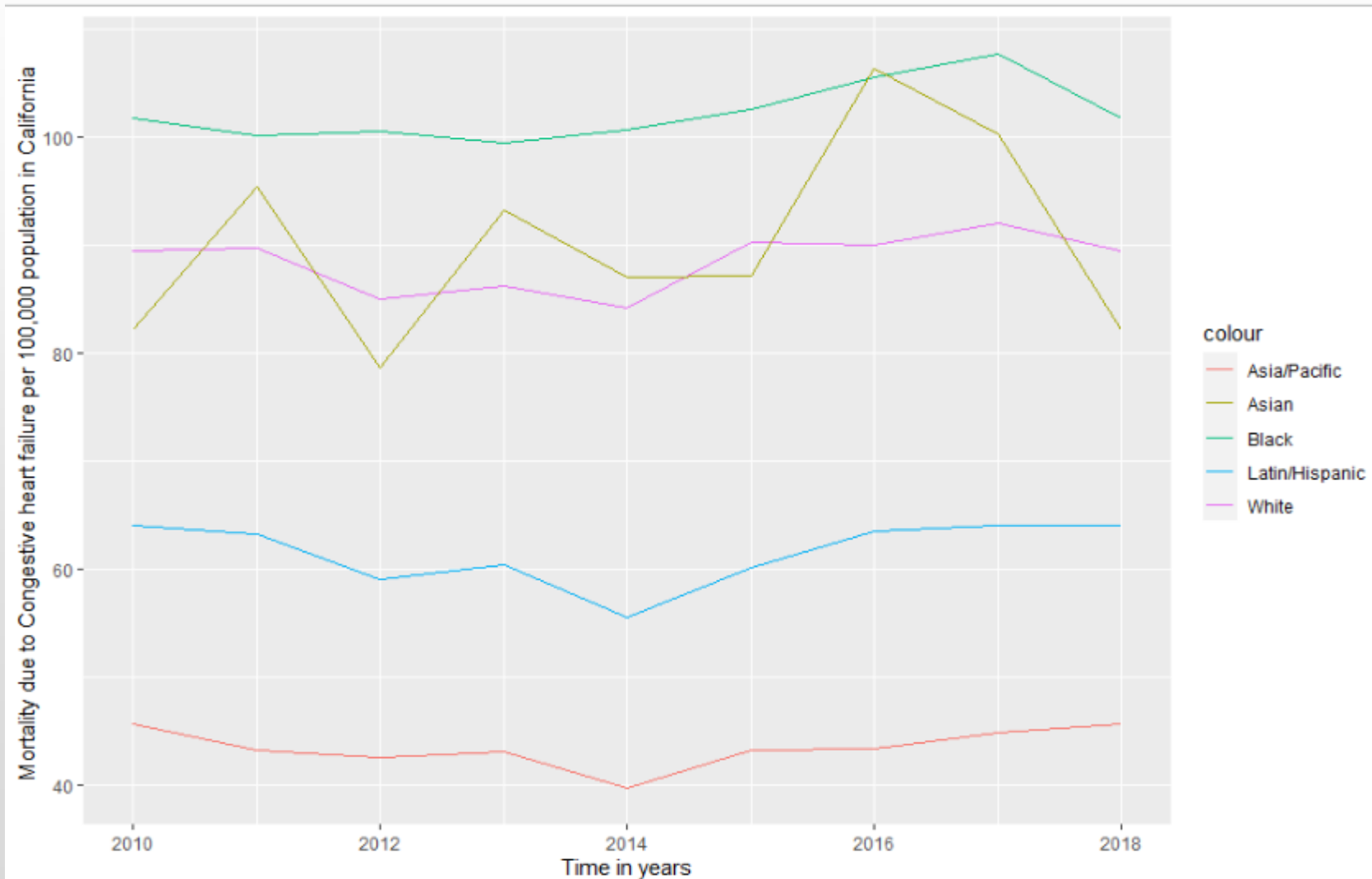
# RACEWISE MORTALITY DUE TO CHF

## New York

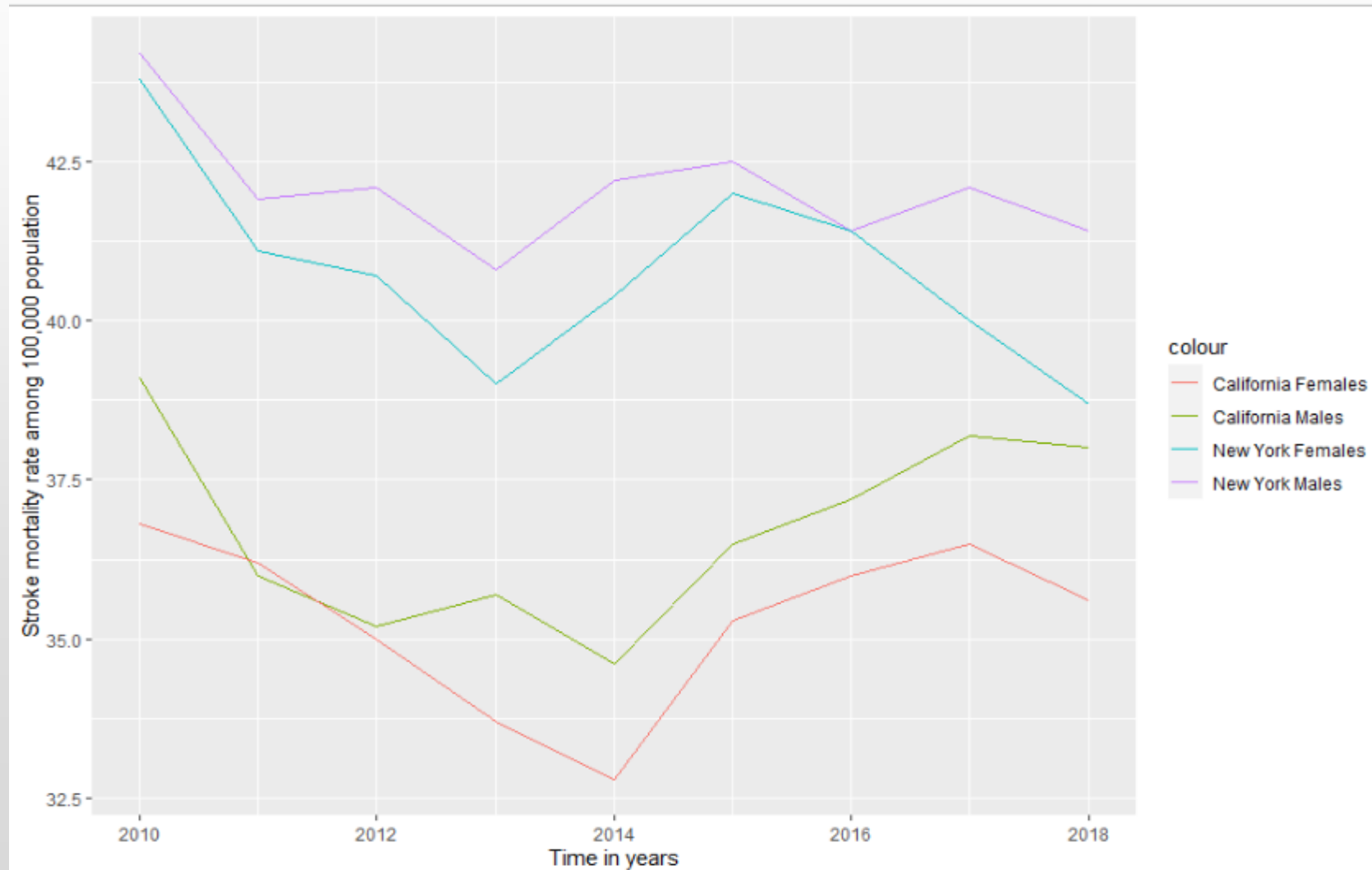


# RACEWISE MORTALITY DUE TO CHF

## California



# ANOMALY IN STROKE / NEW YORK



# Who wins? (*inappropriate framing*)

- Indians have been included in the category of Asians, and American-Indians have been shown to be much more at risk for heart failure when compared to the white population.
- However, the **black** population take the #1 spot at risk for all kinds of heart diseases.

# Racewise SMR of Congestive heart failure

- SMR (averaged for all the years is calculated)
- $= ( \text{HispanicMortality}, \text{WhiteMortality}, \textbf{BlackMortality}, \text{AsianMortality}, \text{PacificMortality} ) \div \text{Overall Mortality}$

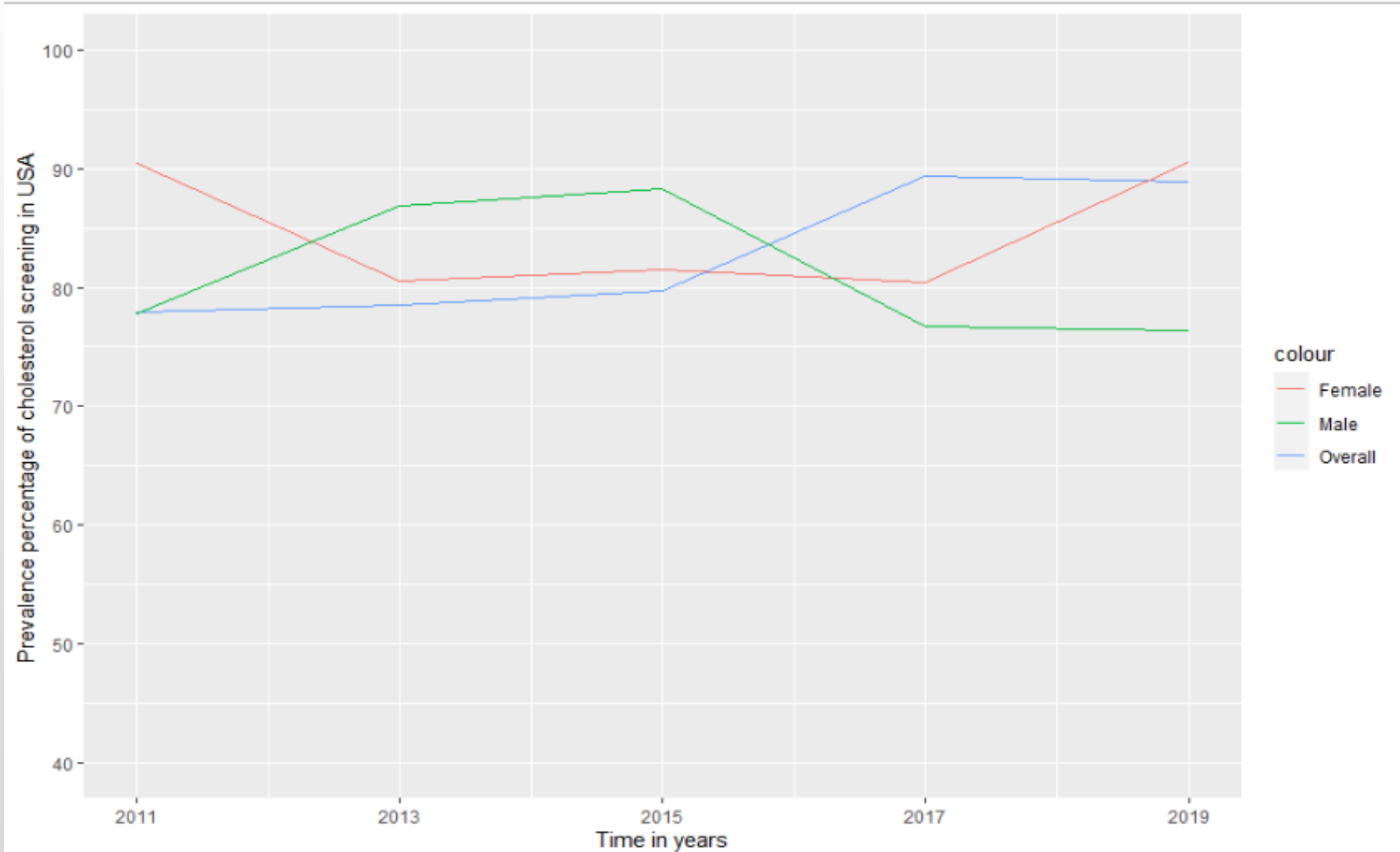
$= (0.78, 1.13, \textbf{1.30}, 1.15, 0.56)$

- Blacks are 1.3 times as likely as the overall population to die of a heart related event caused by congestive heart failure

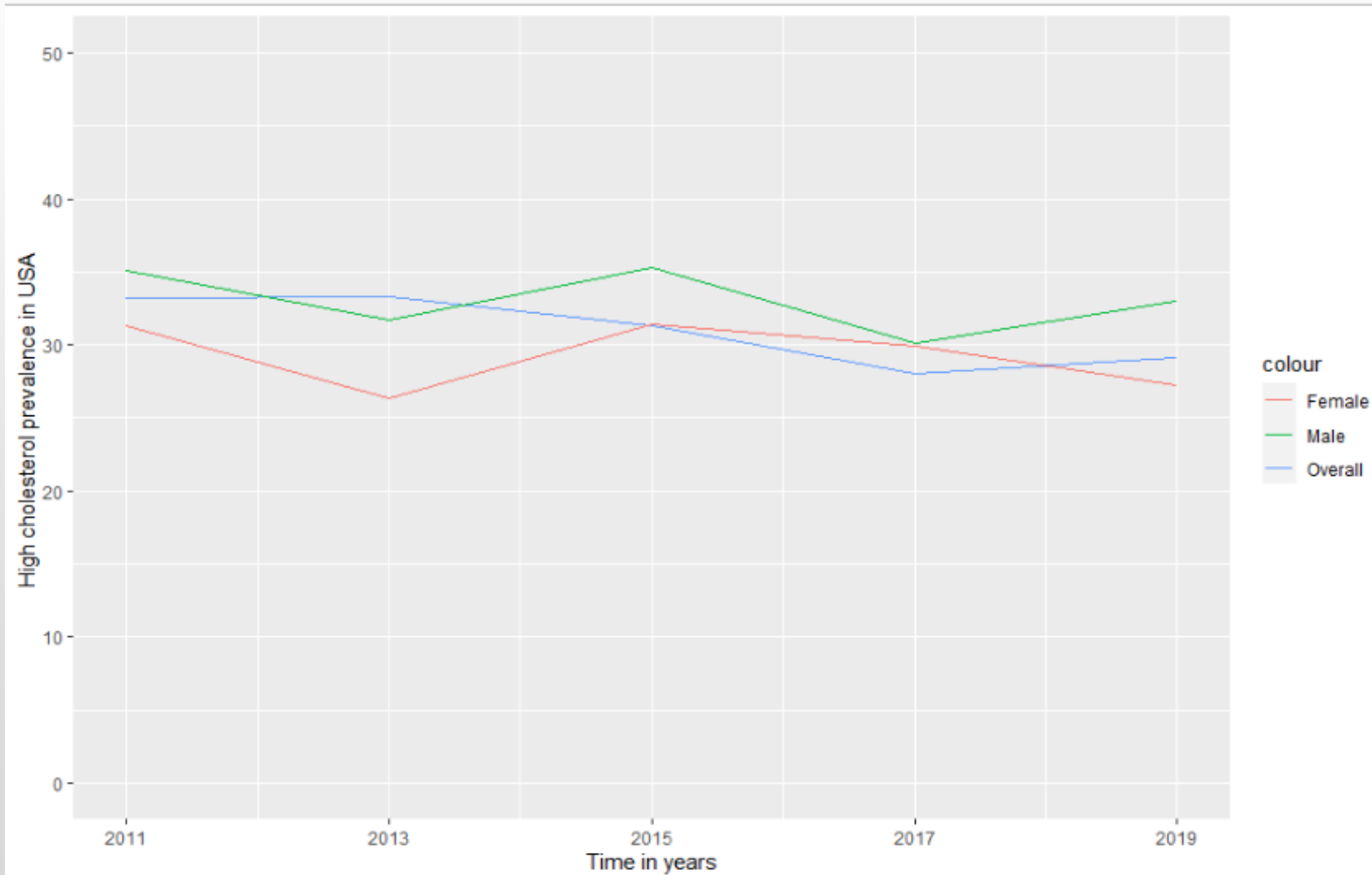


**Q1:** How common is cholesterol screening among adults (18+)? And what is the high cholesterol prevalence ?

## CHOLESTEROL SCREENING PREVALENCE



$$\text{High cholesterol prevalence} = \frac{\text{The number of people at risk screened positive for high cholesterol}}{\text{Total number of people at risk screened}}$$

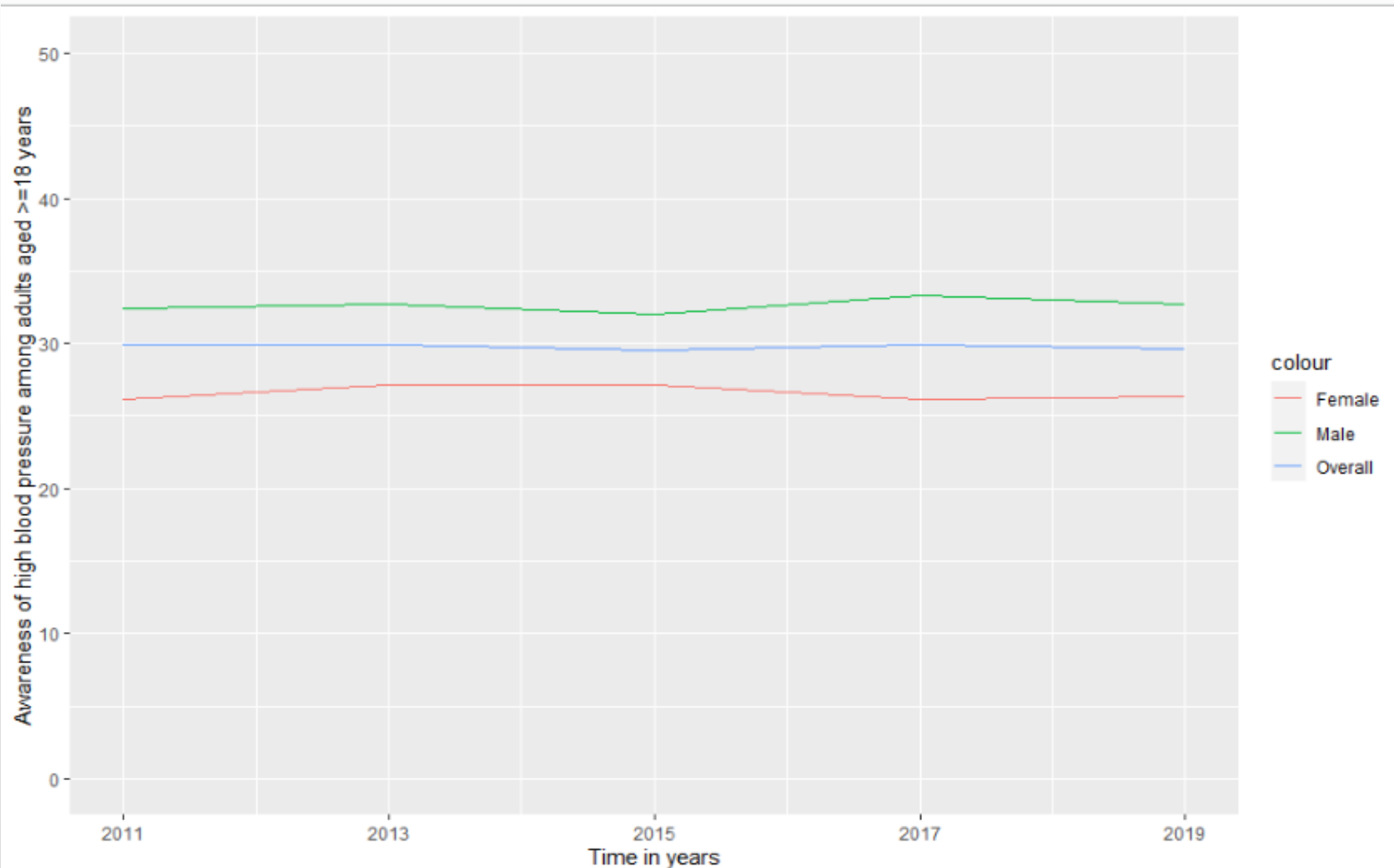


HIGH  
CHOLESTEROL  
PREVALENCE  
AMONG THOSE  
TESTED

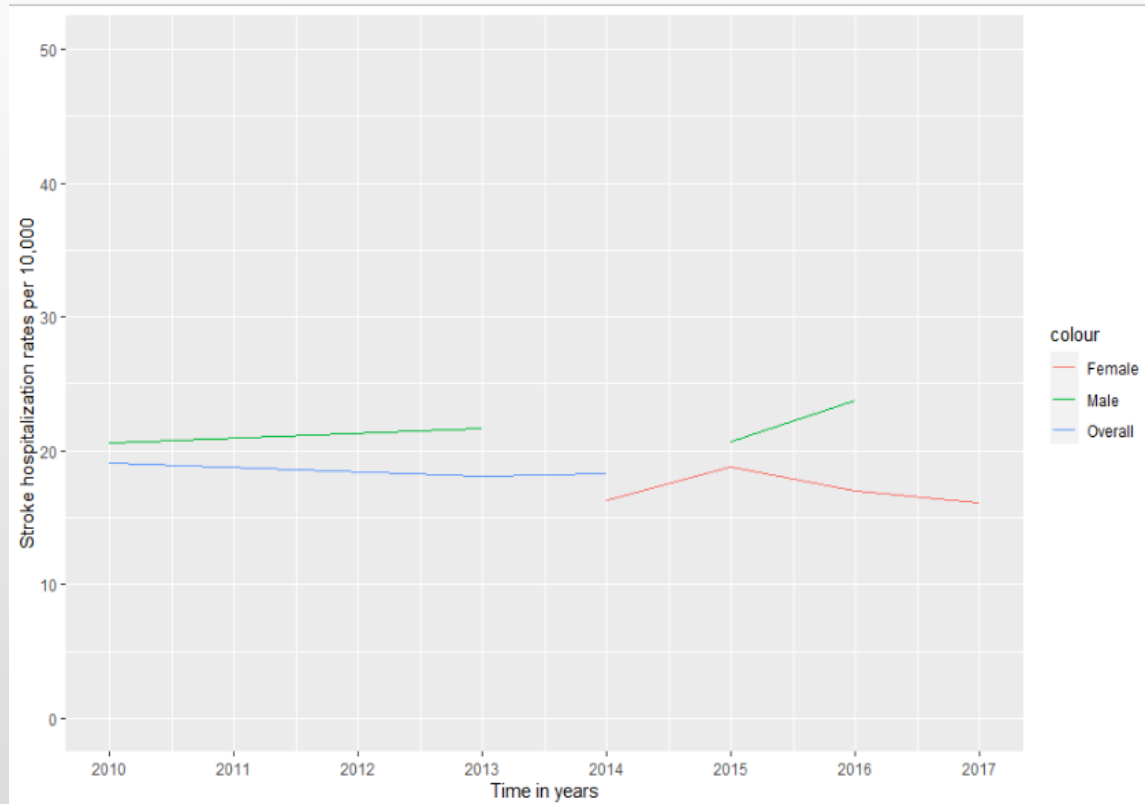
# Inferences

- A **much higher percentage** of the male population could be living with high cholesterol **without getting screened**.
- The minority population (Blacks) could have a **higher mortality** because of inadequate screening

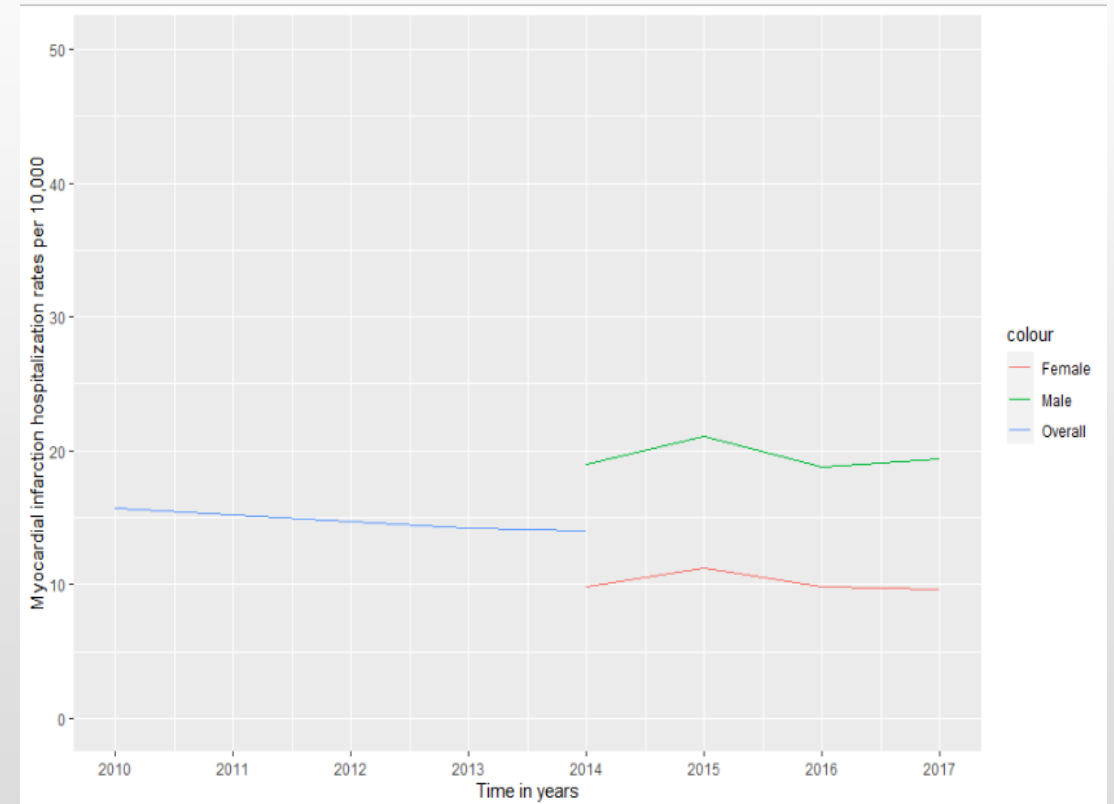
## Q2: What is the prevalence of high blood pressure (hypertension) in the adult population in USA?



# Q3: What is the hospitalization rate for stroke and acute myocardial infarction?



**STROKE**



**MYOCARDIAL INFARCTION**

# How it was done

methods:

- All coding and analysis in R, codes attached here: (we varied one variable to get answers for the rest)
- Code has been attached in a separate .txt file

# Noteworthy points

- The alpha numeric coding of the US states aided us a lot to create time series for all the states of mortality due to all the four conditions
- The coding of the categories as: : { "OVR" "GENF" "GENM" "HIS" "WHT" "BLK" "AIAN" "API" } helped us analyze the data easier by making accessibility better
- There were a lot of “NA” values which led me to restrict the data sets significantly, and there were still further “NA”s in the hospitalization due to stroke and MI.

# What I could salvage

- Created a large list of 11 elements each corresponding to one year from 2010 to 2020 of ONLY data related to cardiovascular disease
- Each list was a large matrix containing 27 types of data entries, including crude rates, age adjusted rates, raw numbers, immunizations, hospitalization rates, prevalence rates etc
- This can be done similarly for all the states, and for all conditions. I have only considered the **Age adjusted** rates to eliminate the differences in the age compositions of the states, and therefore is the more accurate measure when compared to crude rates.



# In hindsight

- We had to consider a US dataset due to the unavailability of an Indian dataset
- India should start making it's hospital data more transparent by **developing newer E-H-R standards** and phasing out the outdated
- While doing so, a standard format (like the aforementioned for alphanumeric coding, coding of categories) should be framed so that the data because readily accessible

# In retrospect

- Need to mitigate creeping up rates of cardiovascular diseases
- Awareness of better diet options, **frequent screening**, more accessible health centres with advanced equipment, and natural primary prevention measures like exercising, reducing salt intake should be popularized
- Studies have shown that Indians are more susceptible to cardiovascular disease than Caucasians, who are the main topic of this analysis
- So, it becomes all the more important for India to step up cardiovascular care facilities, and along with that, better information systems and management

# Looking forward

- Plan to analyse state-wise cancer statistics which is a huge subset of the already big data and try to compare cancer prevalence rates and mortality rates of different states.
- We will also consider the **geographical** aspect (using the **GPS** location given in the data) and try to apply a primitive learning algorithm to determine the “**clustering areas**” of high cancer rates , and will try to draw insights and explain them.

# BIBLIOGRAPHY

- <https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi-d1f3d>
- <https://www.apollohospitals.com/>
- <https://www.who.int/health-topics/cardiovascular-diseases>
- <https://www.nhs.uk/conditions/coronary-heart-disease>
- <https://www.cdc.gov/heartdisease/facts.htm>
- <https://www.healthline.com/health/heart-disease>

Welcome to the final slide. Please go back or go to the next presentation

