Technical challenge

Group 5

2024-08-03

Table of Contents

#Setup for the project

*# Set global chunk options*  
knitr**::**opts\_chunk**$set**(echo = TRUE, error = TRUE, warning = TRUE, message = TRUE)  
*# Set working directory if necessary*  
**getwd**()

## [1] "C:/Users/nkemj/OneDrive/Documents"

# Load packages

**library**(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

**library**(survey)

## Loading required package: grid  
## Loading required package: Matrix  
##   
## Attaching package: 'Matrix'  
##   
## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack  
##   
## Loading required package: survival  
##   
## Attaching package: 'survey'  
##   
## The following object is masked from 'package:graphics':  
##   
## dotchart

*#library(patchwork)*

#Import data

diabetes\_data <- **read.csv**("C:/Users/nkemj/Downloads/Challenge Datasets.csv")  
  
  
**View**(diabetes\_data)  
  
  
*#Recode the variables*  
  
diabetes\_data <- diabetes\_data **%>%**  
 **mutate**(  
 hypertension = **factor**(hypertension, levels = **c**(1, 2, 3, 4, 7, 9), labels = **c**("Yes", "Yes, but female told only during pregnancy", "No", "Told borderline high or pre-hypertensive or elevated blood pressure ", "Don't Know/Not Sure", "Refused")),  
 high\_cholesterol = **factor**(high\_cholesterol, levels= **c**(1, 2, 7, 9), labels = **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 cholesterol\_medication= **factor**(cholesterol\_medication, levels = **c**(1, 2, 7, 9), labels = **c**("Yes", "No", "Don't Know", "Refused")),  
 myocardial\_infarction = **factor**( myocardial\_infarction, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 coronary\_heart\_dx = **factor**(coronary\_heart\_dx, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 asthma = **factor**( asthma, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 skin\_cancer = **factor**( skin\_cancer, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 any\_other\_cancer = **factor**( any\_other\_cancer, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 copd = **factor**( copd, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 kidney\_disease = **factor**( kidney\_disease, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 diabetes\_mellitus = **factor**( diabetes\_mellitus, levels = **c**(1, 2, 3, 4, 7, 9), labels= **c**("Yes", "Yes, but female told only during pregnancy ","No","No, pre-diabetes or borderline diabetes", "Don't Know/Not Sure", "Refused")),  
 arthritis = **factor**( arthritis, levels = **c**(1, 2), labels= **c**("Diagnosed with arthritis ", "Not diagnosed with arthritis")),  
 marital\_status = **factor**(marital\_status, levels = **c**(1, 2, 3, 4, 5, 6, 9), labels = **c**("Married", "Divorced", "Widowed", "Separated", "Never Married", "A member of an unmarried couple", "Refused")),  
 deaf = **factor**( deaf, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 blind = **factor**( blind, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 difficulty\_walking = **factor**( difficulty\_walking, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 hiv\_aids = **factor**( hiv\_aids, levels = **c**(1, 2, 7, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure", "Refused")),  
 caregiver = **factor**( caregiver, levels = **c**(1, 2, 7, 8, 9), labels= **c**("Yes", "No", "Don't Know/Not Sure","Caregiving recipient died in past 30 days", "Refused")),   
 sex = **factor**( sex, levels = **c**(1, 2, 7, 9), labels= **c**("Male", "Female", "Don't Know/Not Sure", "Refused")),  
 sexual\_orientation = **factor**( sexual\_orientation, levels = **c**(1, 2, 3, 4, 7, 9), labels= **c**("Gay", "Straight","Bisexual", "Something else", "Don't Know/Not Sure", "Refused")),  
 ethnicity = **factor**( ethnicity, levels = **c**(1, 2, 3, 4, 5, 6), labels= **c**("White, Non-Hispanic", "Black, Non-Hispanic", "Asian, Non-Hispanic ", "American Indian/Alaskan Native, Non-Hispanic", "Hispanic", "Other race")),  
 physical\_activity = **factor**(physical\_activity, levels = **c**(1, 2, 9), labels = **c**("Had physical activity or exercise ", "No physical activity or exercise in last 30 days", "Don’t know/Refused/Missing")),  
 chd\_or\_mi = **factor**(chd\_or\_mi, levels = **c**(1, 2), labels = **c**("Reported having MI or CHD", "Did not report having MI or CHD")),  
 smoking\_status = **factor**(smoking\_status, levels = **c**(1, 2, 3, 4, 9), labels = **c**("Current smoker - now smokes every day", "Current smoker - now smokes some days", "Former smoker", "Never Smoked", "Don't Know/Refused/Missing"))  
  
)   
  
*# Check fo missing values*  
missing\_values<- **sum**(**is.na**(diabetes\_data))  
missing\_values

## [1] 2126583

*# Check for duplicate rows*  
duplicate\_rows<- **sum**(**duplicated**(diabetes\_data))  
duplicate\_rows

## [1] 0

*# Set options for allowing a single observation per stratum*  
**options**(survey.lonely.psu = "adjust")  
  
*# Create survey design object to account for complex survey design*  
  
data\_design <- **svydesign**(id = **~**1, strata = **~**X\_STSTR, weights = **~**X\_LLCPWT, data = diabetes\_data)

#Data Exploration

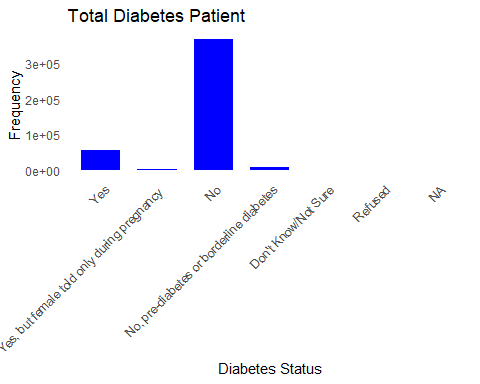
*#Total number of patients with diabetes*  
  
diabetes\_total <- diabetes\_data **%>%**  
 **count**(diabetes\_mellitus) **%>%**  
 **mutate**(percentage = **round**(n **/sum**(n) **\*** 100,0))  
  
**print**(diabetes\_total)

## diabetes\_mellitus n percentage  
## 1 Yes 57616 13  
## 2 Yes, but female told only during pregnancy 3808 1  
## 3 No 366342 84  
## 4 No, pre-diabetes or borderline diabetes 9942 2  
## 5 Don't Know/Not Sure 613 0  
## 6 Refused 369 0  
## 7 <NA> 3 0

*#calculate the diabetes prevalence of the weighted variables*  
weighted\_total <- **svymean**(**~**diabetes\_mellitus, data\_design, na.rm = TRUE)  
  
***##Confidence Interval***  
  
overall\_ci <- **confint**(weighted\_total)  
  
*# Print overall prevalence*  
**cat**("Total diabetes:",   
 **format**(**round**(weighted\_total[1] **\*** 100, 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(overall\_ci[1] **\*** 100, 1), nsmall = 1), "-",  
 **format**(**round**(overall\_ci[2] **\*** 100, 1), nsmall = 1), "%)**\n\n**")

## Total diabetes: 11.4 % (95% CI: 11.1 - 0.9 %)

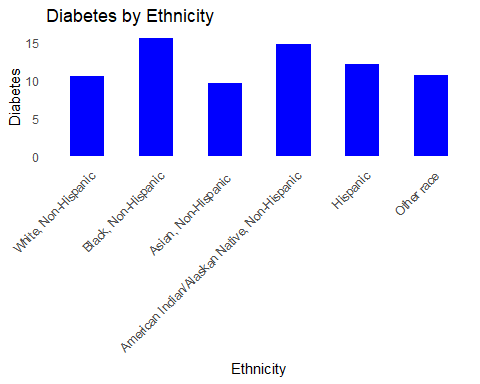
*# Plot the total diabetes*   
diabetes\_total **%>%**  
 **ggplot**(**aes**(diabetes\_mellitus, n)) **+**  
 **geom\_bar**(stat = "identity", fill = "blue", width = 0.7 ) **+**   
 **labs**(  
 title = "Total Diabetes Patient",  
 x = "Diabetes Status",  
 y = "Frequency") **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



*#Diabetes patient by ethnicity*  
  
  
Dia\_eth <- **svyby**(**~**diabetes\_mellitus, **~**ethnicity, data\_design, svymean, na.rm = TRUE)  
  
  
*#calculate Confidence Interval*   
ethnicity\_ci <- **confint**(Dia\_eth)  
  
*#The value of ethnicity\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 6 rows where diabetes = yes.*  
  
ethnicity\_ci\_subset <- ethnicity\_ci[1**:**6,]  
  
*# Combine the results*  
ethnicity\_results <- **data.frame**(  
 Ethnicity = Dia\_eth**$**ethnicity,  
 Diabetes = Dia\_eth**$**diabetes\_mellitusYes **\*** 100,  
 CI\_Lower = ethnicity\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = ethnicity\_ci\_subset[, 2] **\*** 100  
)  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(ethnicity\_results)) {  
 **cat**(ethnicity\_results**$**Ethnicity[i], ":",   
 **format**(**round**(ethnicity\_results**$**Diabetes[i], 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(ethnicity\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(ethnicity\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 10.5 % (95% CI: 10.3 - 10.7 %)  
## 2 : 15.5 % (95% CI: 14.8 - 16.2 %)  
## 3 : 9.6 % (95% CI: 8.2 - 11.0 %)  
## 4 : 14.8 % (95% CI: 13.4 - 16.2 %)  
## 5 : 12.1 % (95% CI: 11.4 - 12.7 %)  
## 6 : 10.7 % (95% CI: 9.6 - 11.9 %)

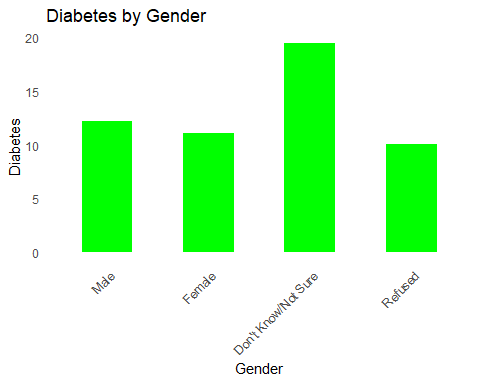
*# Plot Diabetes by ethnicity*   
**ggplot**(ethnicity\_results, **aes**(x = Ethnicity, y = Diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "blue", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes by Ethnicity",  
 x = "Ethnicity",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



*#Diabetes patient by Gender*  
  
  
Gender <- **svyby**(**~**diabetes\_mellitus, **~**sex, data\_design, svymean, na.rm = TRUE)  
  
  
*#calculate Confidence Interval*   
Gender\_ci <- **confint**(Gender)  
  
*#The value of Gender\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 4 rows where diabetes = yes.*  
  
Gender\_ci\_subset <- Gender\_ci[1**:**4,]  
  
*# Combine the results*  
Gender\_results <- **data.frame**(  
 gender = Gender**$**sex,  
 Diabetes = Gender**$**diabetes\_mellitusYes **\*** 100,  
 CI\_Lower = Gender\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = Gender\_ci\_subset[, 2] **\*** 100  
)  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(Gender\_results)) {  
 **cat**(Gender\_results**$**gender[i], ":",   
 **format**(**round**(Gender\_results**$**Diabetes[i], 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(Gender\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(Gender\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 12.2 % (95% CI: 11.7 - 12.7 %)  
## 2 : 11.1 % (95% CI: 10.6 - 11.5 %)  
## 3 : 19.5 % (95% CI: 6.2 - 32.8 %)  
## 4 : 10.1 % (95% CI: 5.0 - 15.2 %)

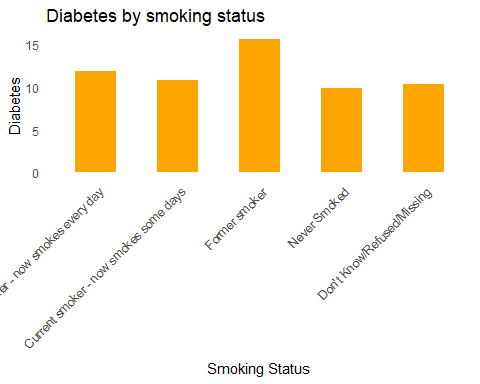
*# Plot Diabetes by ethnicity*   
**ggplot**(Gender\_results, **aes**(x = gender, y = Diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "green", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes by Gender",  
 x = "Gender",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



*#Diabetes patient by Smoking Status*  
  
  
smoking <- **svyby**(**~**diabetes\_mellitus, **~**smoking\_status, data\_design, svymean, na.rm = TRUE)  
  
  
*#calculate Confidence Interval*   
smoking\_ci <- **confint**(smoking)  
  
*#The value of smoking\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 5 rows where diabetes = yes.*  
  
smoking\_ci\_subset <- smoking\_ci[1**:**5,]  
  
*# Combine the results*  
smoking\_results <- **data.frame**(  
 Smoking = smoking**$**smoking\_status,  
 Diabetes = smoking**$**diabetes\_mellitusYes **\*** 100,  
 CI\_Lower = smoking\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = smoking\_ci\_subset[, 2] **\*** 100  
)  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(smoking\_results)) {  
 **cat**(smoking\_results**$**Smoking[i], ":",   
 **format**(**round**(smoking\_results**$**Diabetes[i], 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(smoking\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(smoking\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 11.9 % (95% CI: 11.2 - 12.5 %)  
## 2 : 10.8 % (95% CI: 9.7 - 11.9 %)  
## 3 : 15.6 % (95% CI: 15.1 - 16.0 %)  
## 4 : 9.8 % (95% CI: 9.6 - 10.1 %)  
## 5 : 10.3 % (95% CI: 9.4 - 11.2 %)

*# Plot Diabetes by smoking*  
**ggplot**(smoking\_results, **aes**(x = Smoking, y = Diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "orange", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes by smoking status",  
 x = "Smoking Status",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())

 #Comorbidities

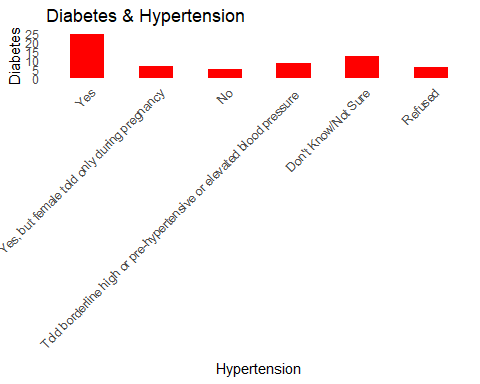
*#Diabetes and hypertension*  
  
Dia\_hyp <- **svyby**(**~**diabetes\_mellitus, **~**hypertension, data\_design, svymean, na.rm = TRUE)  
  
**print**(Dia\_hyp)

## hypertension  
## Yes Yes  
## Yes, but female told only during pregnancy Yes, but female told only during pregnancy  
## No No  
## Told borderline high or pre-hypertensive or elevated blood pressure Told borderline high or pre-hypertensive or elevated blood pressure   
## Don't Know/Not Sure Don't Know/Not Sure  
## Refused Refused  
## diabetes\_mellitusYes  
## Yes 0.24411949  
## Yes, but female told only during pregnancy 0.06490515  
## No 0.04924415  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.08364222  
## Don't Know/Not Sure 0.12029725  
## Refused 0.06197352  
## diabetes\_mellitusYes, but female told only during pregnancy   
## Yes 0.006584214  
## Yes, but female told only during pregnancy 0.097438007  
## No 0.010393223  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.008576644  
## Don't Know/Not Sure 0.008645223  
## Refused 0.002994835  
## diabetes\_mellitusNo  
## Yes 0.7097617  
## Yes, but female told only during pregnancy 0.8160211  
## No 0.9228545  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.8060398  
## Don't Know/Not Sure 0.7429432  
## Refused 0.6655488  
## diabetes\_mellitusNo, pre-diabetes or borderline diabetes  
## Yes 0.03734821  
## Yes, but female told only during pregnancy 0.02103858  
## No 0.01594585  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.10059256  
## Don't Know/Not Sure 0.09622121  
## Refused 0.01997729  
## diabetes\_mellitusDon't Know/Not Sure  
## Yes 0.0018675986  
## Yes, but female told only during pregnancy 0.0002139258  
## No 0.0012237493  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.0011487266  
## Don't Know/Not Sure 0.0267294487  
## Refused 0.0141285427  
## diabetes\_mellitusRefused  
## Yes 0.0003187959  
## Yes, but female told only during pregnancy 0.0003832305  
## No 0.0003385447  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.0000000000  
## Don't Know/Not Sure 0.0051636861  
## Refused 0.2353769876  
## se.diabetes\_mellitusYes  
## Yes 0.0023857390  
## Yes, but female told only during pregnancy 0.0129935190  
## No 0.0009479911  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.0113090264  
## Don't Know/Not Sure 0.0244673753  
## Refused 0.0121850958  
## se.diabetes\_mellitusYes, but female told only during pregnancy   
## Yes 0.0003624858  
## Yes, but female told only during pregnancy 0.0106453452  
## No 0.0004969470  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.0023559863  
## Don't Know/Not Sure 0.0038467529  
## Refused 0.0016884117  
## se.diabetes\_mellitusNo  
## Yes 0.002546150  
## Yes, but female told only during pregnancy 0.016041841  
## No 0.001220521  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.016287600  
## Don't Know/Not Sure 0.053501180  
## Refused 0.034503358  
## se.diabetes\_mellitusNo, pre-diabetes or borderline diabetes  
## Yes 0.0012257987  
## Yes, but female told only during pregnancy 0.0049201546  
## No 0.0006342104  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.0133081970  
## Don't Know/Not Sure 0.0569453634  
## Refused 0.0107550140  
## se.diabetes\_mellitusDon't Know/Not Sure  
## Yes 0.0001989126  
## Yes, but female told only during pregnancy 0.0001443621  
## No 0.0001381365  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.0008096460  
## Don't Know/Not Sure 0.0075247096  
## Refused 0.0081898394  
## se.diabetes\_mellitusRefused  
## Yes 7.118387e-05  
## Yes, but female told only during pregnancy 2.765838e-04  
## No 6.021692e-05  
## Told borderline high or pre-hypertensive or elevated blood pressure 0.000000e+00  
## Don't Know/Not Sure 3.048990e-03  
## Refused 2.948478e-02

*#calculate Confidence Interval*   
dia\_hyp\_ci <- **confint**(Dia\_hyp)  
  
*#The value of dia\_hyp\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 6 rows where diabetes = yes.*  
  
dia\_hyp\_ci\_subset <- dia\_hyp\_ci[1**:**6,]  
  
*# Combine the results*  
dia\_hyp\_results <- **data.frame**(  
 Hypertension = Dia\_hyp**$**hypertension,  
 Diabetes = Dia\_hyp**$**diabetes\_mellitusYes **\*** 100,  
 CI\_Lower = dia\_hyp\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = dia\_hyp\_ci\_subset[, 2] **\*** 100  
)  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(dia\_hyp\_results)) {  
 **cat**(dia\_hyp\_results**$**Hypertension[i], ":",   
 **format**(**round**(dia\_hyp\_results**$**Diabetes[i], 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(dia\_hyp\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(dia\_hyp\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 24.4 % (95% CI: 23.9 - 24.9 %)  
## 2 : 6.5 % (95% CI: 3.9 - 9.0 %)  
## 3 : 4.9 % (95% CI: 4.7 - 5.1 %)  
## 4 : 8.4 % (95% CI: 6.1 - 10.6 %)  
## 5 : 12.0 % (95% CI: 7.2 - 16.8 %)  
## 6 : 6.2 % (95% CI: 3.8 - 8.6 %)

*# Plot Diabetes by hypertension*   
**ggplot**(dia\_hyp\_results, **aes**(x = Hypertension, y = Diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "red", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes & Hypertension",  
 x = "Hypertension",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



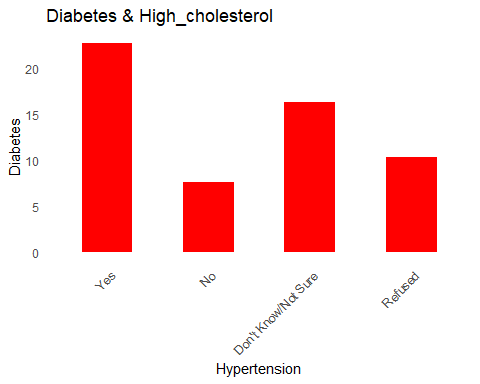
*#Diabetes and High cholesterol*  
  
Dia\_chol <- **svyby**(**~**diabetes\_mellitus, **~**high\_cholesterol, data\_design, svymean, na.rm = TRUE)  
  
  
*#calculate Confidence Interval*   
dia\_chol\_ci <- **confint**(Dia\_chol)  
  
*#The value of dia\_chol\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 6 rows where diabetes = yes.*  
  
dia\_chol\_ci\_subset <- dia\_chol\_ci[1**:**4,]  
  
dia\_chol\_ci\_subset2 <- dia\_chol\_ci[5**:**8,]  
  
*# Combine the results*  
dia\_chol\_results <- **data.frame**(  
 High\_cholesterol = Dia\_chol**$**high\_cholesterol,  
 Diabetes = Dia\_chol**$**diabetes\_mellitusYes **\*** 100,  
 Diabetes2 = Dia\_chol**$**`diabetes\_mellitusYes, but female told only during pregnancy `**\*** 100,  
 CI\_Lower = dia\_chol\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = dia\_chol\_ci\_subset[, 2] **\*** 100  
)  
  
  
*# Combine the results for diabetes 2*  
dia\_chol\_results2 <- **data.frame**(  
 High\_cholesterol = Dia\_chol**$**high\_cholesterol,  
 Diabetes2 = Dia\_chol**$**`diabetes\_mellitusYes, but female told only during pregnancy `**\*** 100,  
 CI\_Lower = dia\_chol\_ci\_subset2[, 1] **\*** 100,  
 CI\_Upper = dia\_chol\_ci\_subset2[, 2] **\*** 100  
)  
  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(dia\_chol\_results)) {  
 **cat**(dia\_chol\_results**$**High\_cholesterol[i], ":",   
 **format**(**round**(dia\_chol\_results**$**Diabetes[i], 1), nsmall = 1),  
 "%",  
 "(95% CI:",   
 **format**(**round**(dia\_chol\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(dia\_chol\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 22.7 % (95% CI: 22.2 - 23.2 %)  
## 2 : 7.6 % (95% CI: 7.4 - 7.9 %)  
## 3 : 16.3 % (95% CI: 13.3 - 19.3 %)  
## 4 : 10.3 % (95% CI: 5.3 - 15.2 %)

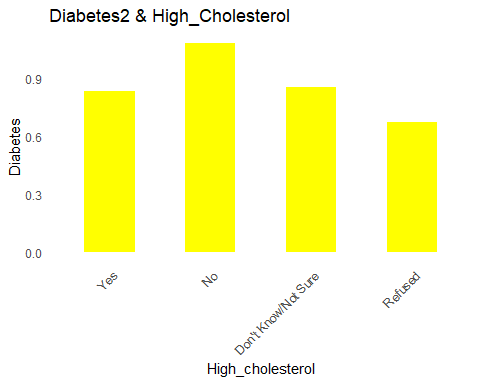
*#print results for diabetes 2*  
**for**(i **in** 1**:nrow**(dia\_chol\_results2)) {  
 **cat**(dia\_chol\_results2**$**High\_cholesterol[i], ":",   
 **format**(**round**(dia\_chol\_results2**$**Diabetes2[i], 1), nsmall = 1),  
 "%",  
 "(95% CI:",   
 **format**(**round**(dia\_chol\_results2**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(dia\_chol\_results2**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 0.8 % (95% CI: 0.7 - 1.0 %)  
## 2 : 1.1 % (95% CI: 1.0 - 1.2 %)  
## 3 : 0.9 % (95% CI: 0.3 - 1.4 %)  
## 4 : 0.7 % (95% CI: 0.0 - 1.4 %)

*# Plot Diabetes by High\_cholesterol*  
 **ggplot**(dia\_chol\_results, **aes**(x = High\_cholesterol, y = Diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "red", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes & High\_cholesterol",  
 x = "Hypertension",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



**ggplot**(dia\_chol\_results, **aes**(x = High\_cholesterol, y = Diabetes2)) **+**  
 **geom\_bar**(stat = "identity", fill = "yellow", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes2 & High\_Cholesterol",  
 x = "High\_cholesterol",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



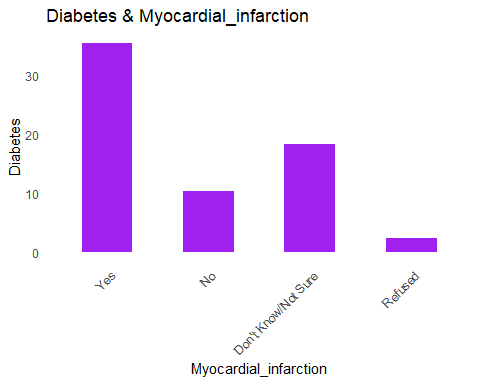
*# Display both charts side by side with the patchwork package*  
*#p1 + p2*

#Comorbidities

*#Diabetes and myocardial\_infarction*  
  
Dia\_myo <- **svyby**(**~**diabetes\_mellitus, **~**myocardial\_infarction, data\_design, svymean, na.rm = TRUE)  
  
  
*#calculate Confidence Interval*   
dia\_myo\_ci <- **confint**(Dia\_myo)  
  
*#The value of dia\_myo\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 6 rows where diabetes = yes.*  
  
dia\_myo\_ci\_subset <- dia\_myo\_ci[1**:**4,]  
  
*# Combine the results*  
dia\_myo\_results <- **data.frame**(  
 Myocardial\_infarction = Dia\_myo**$**myocardial\_infarction,  
 diabetes = Dia\_myo**$**diabetes\_mellitusYes **\*** 100,  
 diabetes2 = Dia\_myo**$**`diabetes\_mellitusYes, but female told only during pregnancy `**\***100,  
 CI\_Lower = dia\_myo\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = dia\_myo\_ci\_subset[, 2] **\*** 100  
)  
  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(dia\_myo\_results)) {  
 **cat**(dia\_myo\_results**$**Myocardial\_infarction[i], ":",   
 **format**(**round**(dia\_myo\_results**$**diabetes[i], 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(dia\_myo\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(dia\_myo\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 35.4 % (95% CI: 34.0 - 36.8 %)  
## 2 : 10.3 % (95% CI: 10.1 - 10.5 %)  
## 3 : 18.3 % (95% CI: 15.1 - 21.4 %)  
## 4 : 2.4 % (95% CI: 1.0 - 3.8 %)

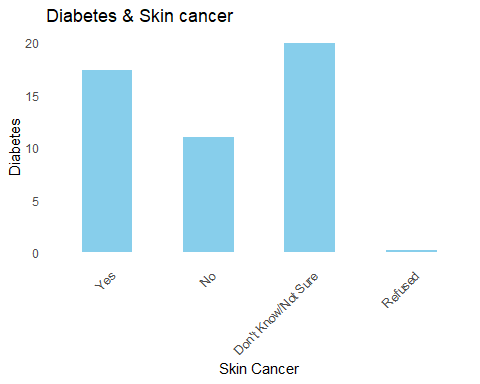
*# Plot Diabetes by Myocardinal*   
**ggplot**(dia\_myo\_results, **aes**(x = Myocardial\_infarction, y = diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "purple", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes & Myocardial\_infarction",  
 x = "Myocardial\_infarction",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



*#Diabetes and Skin cancer*  
  
Dia\_cancer <- **svyby**(**~**diabetes\_mellitus, **~**skin\_cancer, data\_design, svymean, na.rm = TRUE)  
  
  
*#calculate Confidence Interval*   
dia\_cancer\_ci <- **confint**(Dia\_cancer)  
  
*#The value of dia\_cancer\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 6 rows where diabetes = yes.*  
  
dia\_cancer\_ci\_subset <- dia\_cancer\_ci[1**:**4,]  
  
*# Combine the results*  
dia\_cancer\_results <- **data.frame**(  
 Skin\_cancer = Dia\_cancer**$**skin\_cancer,  
 diabetes = Dia\_cancer**$**diabetes\_mellitusYes **\*** 100,  
 diabetes2 = Dia\_cancer**$**`diabetes\_mellitusYes, but female told only during pregnancy `**\***100,  
 CI\_Lower = dia\_cancer\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = dia\_cancer\_ci\_subset[, 2] **\*** 100  
)  
  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(dia\_cancer\_results)) {  
 **cat**(dia\_cancer\_results**$**Skin\_cancer[i], ":",   
 **format**(**round**(dia\_cancer\_results**$**diabetes[i], 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(dia\_cancer\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(dia\_cancer\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 17.3 % (95% CI: 16.5 - 18.1 %)  
## 2 : 11.0 % (95% CI: 10.7 - 11.2 %)  
## 3 : 20.0 % (95% CI: 12.4 - 27.6 %)  
## 4 : 0.2 % (95% CI: -0.1 - 0.6 %)

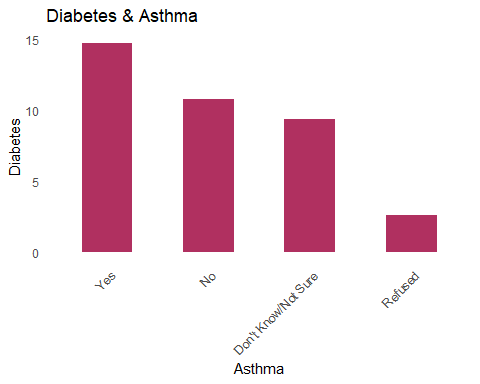
*# Plot Diabetes by Skincancer*   
**ggplot**(dia\_cancer\_results, **aes**(x = Skin\_cancer, y = diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "skyblue", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes & Skin cancer",  
 x = "Skin Cancer",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



*#Diabetes and asthma*  
  
Dia\_asthma <- **svyby**(**~**diabetes\_mellitus, **~**asthma, data\_design, svymean, na.rm = TRUE)  
  
  
*#calculate Confidence Interval*   
dia\_asthma\_ci <- **confint**(Dia\_asthma)  
  
*#The value of dia\_asthma\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 4 rows where diabetes = yes.*  
  
dia\_asthma\_ci\_subset <- dia\_asthma\_ci[1**:**4,]  
  
*# Combine the results*  
dia\_asthma\_results <- **data.frame**(  
 Asthma = Dia\_asthma**$**asthma,  
 diabetes = Dia\_asthma**$**diabetes\_mellitusYes **\*** 100,  
 diabetes2 = Dia\_asthma**$**`diabetes\_mellitusYes, but female told only during pregnancy `**\***100,  
 CI\_Lower = dia\_asthma\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = dia\_asthma\_ci\_subset[, 2] **\*** 100  
)  
  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(dia\_asthma\_results)) {  
 **cat**(dia\_asthma\_results**$**Asthma[i], ":",   
 **format**(**round**(dia\_asthma\_results**$**diabetes[i], 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(dia\_asthma\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(dia\_asthma\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 14.8 % (95% CI: 14.2 - 15.4 %)  
## 2 : 10.8 % (95% CI: 10.6 - 11.0 %)  
## 3 : 9.4 % (95% CI: 7.2 - 11.6 %)  
## 4 : 2.6 % (95% CI: -0.2 - 5.4 %)

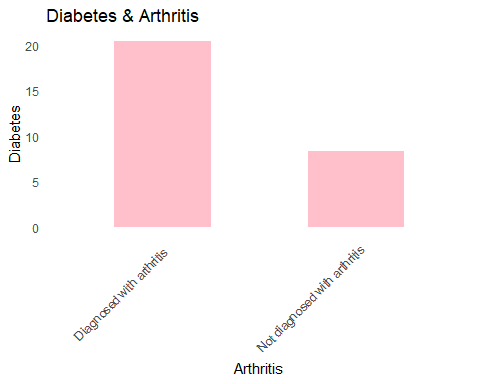
*# Plot Diabetes by Asthma*   
**ggplot**(dia\_asthma\_results, **aes**(x = Asthma, y = diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "maroon", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes & Asthma",  
 x = "Asthma",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



*#Diabetes and Arthritis*  
  
Dia\_art <- **svyby**(**~**diabetes\_mellitus, **~**arthritis, data\_design, svymean, na.rm = TRUE)  
  
  
*#calculate Confidence Interval*   
dia\_art\_ci <- **confint**(Dia\_art)  
  
*#The value of dia\_art\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 4 rows where diabetes = yes.*  
  
dia\_art\_ci\_subset <- dia\_art\_ci[1**:**2,]  
  
*# Combine the results*  
dia\_art\_results <- **data.frame**(  
 Arthritis = Dia\_art**$**arthritis,  
 diabetes = Dia\_art**$**diabetes\_mellitusYes **\*** 100,  
 diabetes2 = Dia\_art**$**`diabetes\_mellitusYes, but female told only during pregnancy `**\***100,  
 CI\_Lower = dia\_art\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = dia\_art\_ci\_subset[, 2] **\*** 100  
)  
  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(dia\_art\_results)) {  
 **cat**(dia\_art\_results**$**Arthritis[i], ":",   
 **format**(**round**(dia\_art\_results**$**diabetes[i], 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(dia\_art\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(dia\_art\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 20.3 % (95% CI: 19.9 - 20.8 %)  
## 2 : 8.3 % (95% CI: 8.1 - 8.5 %)

*# Plot Diabetes by Arthritis*   
**ggplot**(dia\_art\_results, **aes**(x = Arthritis, y = diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "pink", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes & Arthritis",  
 x = "Arthritis",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())



*#Diabetes and HIV/AIDS*  
  
Dia\_hiv <- **svyby**(**~**diabetes\_mellitus, **~**hiv\_aids, data\_design, svymean, na.rm = TRUE)  
  
  
*#calculate Confidence Interval*   
dia\_hiv\_ci <- **confint**(Dia\_hiv)  
  
*#The value of dia\_hiv\_ci gives the result of all the outcome variables to get the exact prevalence, we created a subset of the first 4 rows where diabetes = yes.*  
  
dia\_hiv\_ci\_subset <- dia\_hiv\_ci[1**:**4,]  
  
*# Combine the results*  
dia\_hiv\_results <- **data.frame**(  
 HIV\_AIDS = Dia\_hiv**$**hiv\_aids,  
 diabetes = Dia\_hiv**$**diabetes\_mellitusYes **\*** 100,  
 diabetes2 = Dia\_hiv**$**`diabetes\_mellitusYes, but female told only during pregnancy `**\***100,  
 CI\_Lower = dia\_hiv\_ci\_subset[, 1] **\*** 100,  
 CI\_Upper = dia\_hiv\_ci\_subset[, 2] **\*** 100  
)  
  
  
  
*# Print results*   
**for**(i **in** 1**:nrow**(dia\_hiv\_results)) {  
 **cat**(dia\_hiv\_results**$**HIV\_AIDS[i], ":",   
 **format**(**round**(dia\_hiv\_results**$**diabetes[i], 1), nsmall = 1), "%",  
 "(95% CI:",   
 **format**(**round**(dia\_hiv\_results**$**CI\_Lower[i], 1), nsmall = 1), "-",  
 **format**(**round**(dia\_hiv\_results**$**CI\_Upper[i], 1), nsmall = 1), "%)**\n**")  
}

## 1 : 10.5 % (95% CI: 10.1 - 10.8 %)  
## 2 : 12.0 % (95% CI: 11.7 - 12.3 %)  
## 3 : 12.8 % (95% CI: 11.8 - 13.8 %)  
## 4 : 5.3 % (95% CI: 3.6 - 7.0 %)

*# Plot Diabetes by HIV/AIDS*   
**ggplot**(dia\_hiv\_results, **aes**(x = HIV\_AIDS, y = diabetes)) **+**  
 **geom\_bar**(stat = "identity", fill = "red", width = 0.5) **+**  
 **labs**(  
 title = "Diabetes & HIV/AIDS",  
 x = "HIV/AIDS",  
 y = "Diabetes"  
 ) **+**  
 **theme\_minimal**() **+**  
 **theme**(axis.text.x = **element\_text**(angle = 45, hjust = 1)) **+**   
 **theme**(panel.grid.major = **element\_blank**(),  
 panel.grid.minor = **element\_blank**())

