Final Project

Jennifer Frediani

3/4/2020

## Warning: package 'dplyr' was built under R version 3.6.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Warning: package 'haven' was built under R version 3.6.2

## Warning: package 'tableone' was built under R version 3.6.3

These data are from the Health Information National Trends Survey (HINTS)data from the National Cancer Institute. I chose HINTS 4 Cycle 2 as a starting point for these analyses because it had questions on BMI, cancer history, questions pertaining to screening history and information on diet and exercise. I am interested in exploring relationships between healthy lifestyle behaviors and incidence of prostate cancer in people with and without a family history.

This loaded dataset was simplified in SPSS before importing to R.

Missing data are coded in a variety of ways in this dataset. Most ways are with a negative number of some kind. Eventually we will need to calculate whether participants meet the physical activity guidelines, which will require coding them as something that will be ignored when those are done.

## # A tibble: 3,630 x 3  
## ExerciseModMn0 ExerciseModHr0 ExerciseModDays0  
## <dbl> <dbl> <dbl>  
## 1 0 1 5  
## 2 NA 3 5  
## 3 40 NA 3  
## 4 60 NA 6  
## 5 NA 1 2  
## 6 NA 3 1  
## 7 NA NA NA  
## 8 30 NA 4  
## 9 NA NA NA  
## 10 NA 10 7  
## # ... with 3,620 more rows

## [1] "tbl\_df" "tbl" "data.frame"

## [1] 3630 171

There are 3630 observations in this dataset. For now we will limit the dataset to only those that have or had prostate cancer.

## [1] 60

There are 60 participants that have or had prostate cancer. We’ll now look at of those 60 who has a family history of cancer.

## [1] 50

Well, the majority of those who have/had prostate cancer also had a family history of cancer. So let’s look at healthy behaviors in those with prostate cancer.

## [1] 19

## [1] 9

So we have poor fruit and vegetable intake, not suprising, how are they doing with exercise…

## [1] 23

Not many…

Now let’s look at those with a family history of cancer but have not had cancer yet.

## [1] 2050

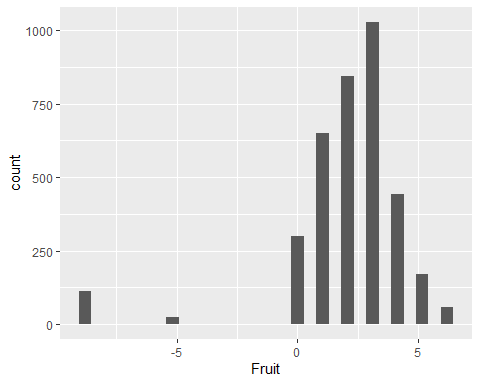
## [1] 604

## [1] 339

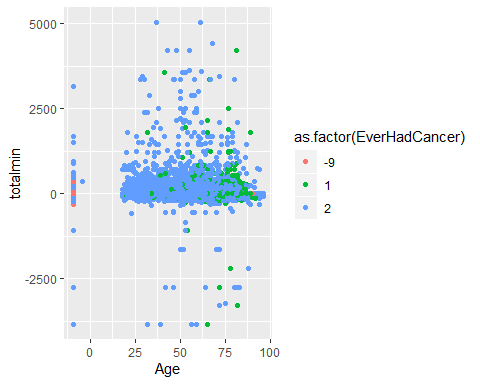
## [1] 598

## Don't know how to automatically pick scale for object of type haven\_labelled. Defaulting to continuous.

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



## Don't know how to automatically pick scale for object of type haven\_labelled. Defaulting to continuous.  
## Don't know how to automatically pick scale for object of type haven\_labelled. Defaulting to continuous.

 Now I’ll create a Table 1 with all of the demographics I think I may use

## # A tibble: 6 x 171  
## HHID PersonID Stratum HISP\_HH RUC2003 CENSDIV CENSREG FormType  
## <dbl> <chr> <chr+lb> <dbl+l> <dbl+l> <dbl+l> <dbl+l> <dbl+lb>  
## 1 2.00e7 2000003~ HM [Hig~ 2 [No] 1 [Cou~ 5 [Sou~ 3 [Sou~ 2 [Long~  
## 2 2.00e7 2000004~ LM [Low~ 2 [No] 6 [Non~ 2 [Mid~ 1 [Nor~ 2 [Long~  
## 3 2.00e7 2000008~ HM [Hig~ 2 [No] 1 [Cou~ 3 [Eas~ 2 [Mid~ 2 [Long~  
## 4 2.00e7 2000010~ HM [Hig~ 2 [No] 1 [Cou~ 5 [Sou~ 3 [Sou~ 2 [Long~  
## 5 2.00e7 2000016~ HM [Hig~ 2 [No] 4 [Non~ 5 [Sou~ 3 [Sou~ 2 [Long~  
## 6 2.00e7 2000018~ HM [Hig~ 2 [No] 1 [Cou~ 5 [Sou~ 3 [Sou~ 2 [Long~  
## # ... with 163 more variables: Language\_Flag <dbl+lbl>, QDisp <dbl+lbl>,  
## # RegularProvider <dbl+lbl>, HealthInsurance <dbl+lbl>,  
## # MostRecentCheckup <dbl+lbl>, Behaviors\_Diabetes <dbl+lbl>,  
## # Behaviors\_Obesity <dbl+lbl>, Behaviors\_HeartDisease <dbl+lbl>,  
## # Behaviors\_HighBP <dbl+lbl>, Behaviors\_Cancer <dbl+lbl>,  
## # Genetics\_Diabetes <dbl+lbl>, Genetics\_Obesity <dbl+lbl>,  
## # Genetics\_HeartDisease <dbl+lbl>, Genetics\_HighBP <dbl+lbl>,  
## # Genetics\_Cancer <dbl+lbl>, GeneralHealth <dbl+lbl>,  
## # OwnAbilityTakeCareHealth <dbl+lbl>, MedConditions\_Diabetes <dbl+lbl>,  
## # MedConditions\_HighBP <dbl+lbl>, MedConditions\_HeartCondition <dbl+lbl>,  
## # MedConditions\_LungDisease <dbl+lbl>, MedConditions\_Arthritis <dbl+lbl>,  
## # MedConditions\_Depression <dbl+lbl>, SleepWorkdayHr <dbl+lbl>,  
## # SleepWorkdayMn <dbl+lbl>, SleepWeekendHr <dbl+lbl>,  
## # SleepWeekendMn <dbl+lbl>, Height\_Feet <dbl+lbl>, Height\_Inches <dbl+lbl>,  
## # Weight <dbl+lbl>, WeightIntention <dbl+lbl>, WeightOpinion <dbl+lbl>,  
## # LittleInterest <dbl+lbl>, Hopeless <dbl+lbl>, Nervous <dbl+lbl>,  
## # Worrying <dbl+lbl>, EmotionalSupport <dbl+lbl>,  
## # UseMenuCalorieInfo <dbl+lbl>, HelpfulMenuCalorieInfo <dbl+lbl>,  
## # Fruit <dbl+lbl>, FruitIntent <dbl+lbl>, Vegetables <dbl+lbl>,  
## # VegetablesIntent <dbl+lbl>, RegularSodaWeek <dbl+lbl>,  
## # RegularSodaIntention <dbl+lbl>, TimesModerateExercise <dbl+lbl>,  
## # HowLongModerateExerciseMn <dbl+lbl>, HowLongModerateExerciseHr <dbl+lbl>,  
## # TimesStrengthTraining <dbl+lbl>, ExerciseIntention <dbl+lbl>,  
## # RegExercise\_Pressure <dbl+lbl>, RegExercise\_Appearance <dbl+lbl>,  
## # RegExercise\_Guilt <dbl+lbl>, RegExercise\_Enjoyment <dbl+lbl>,  
## # AverageDailyTVGames <dbl+lbl>, Smoke100 <dbl+lbl>, SmokeNow <dbl+lbl>,  
## # TriedQuit <dbl+lbl>, ConsiderQuit <dbl+lbl>, GenderC <dbl+lbl>,  
## # DrTalkMammogram <dbl+lbl>, WhenMammogram <dbl+lbl>,  
## # EverHadPSATest <dbl+lbl>, DrShouldPSATest <dbl+lbl>,  
## # DrWantedPSATest <dbl+lbl>, SomeDisagreePSATests <dbl+lbl>,  
## # ProstateCa\_PSATest <dbl+lbl>, ProstateCa\_SlowGrowing <dbl+lbl>,  
## # ProstateCa\_SideEffects <dbl+lbl>, EverHadCancer <dbl+lbl>,  
## # CaBreast <dbl+lbl>, CaProstate <dbl+lbl>, Cancer\_Cat <dbl+lbl>,  
## # WhenDiagnosedCancer <dbl+lbl>, UndergoCancerTreatment <dbl+lbl>,  
## # CancerTx\_Chemo <dbl+lbl>, CancerTx\_Radiation <dbl+lbl>,  
## # CancerTx\_Surgery <dbl+lbl>, CancerTx\_Other <dbl+lbl>,  
## # HowLongFinishTreatment\_Cat <dbl+lbl>, CancerTxSummary <dbl+lbl>,  
## # CancerCheckups <dbl+lbl>, FamilyEverHadCancer <dbl+lbl>, Age <dbl+lbl>,  
## # OccupationStatus <dbl+lbl>, OccupationStatus\_OS <chr+lbl>,  
## # Employed <dbl+lbl>, Unemployed <dbl+lbl>, Homemaker <dbl+lbl>,  
## # Student <dbl+lbl>, Retired <dbl+lbl>, Disabled <dbl+lbl>,  
## # OtherOcc <dbl+lbl>, MultiOcc <dbl>, MaritalStatus <dbl+lbl>,  
## # Education <dbl+lbl>, BornInUSA <dbl+lbl>, YearCameToUSA <dbl+lbl>,  
## # NotHisp <dbl+lbl>, Hisp\_Cat <dbl+lbl>, ...

## Warning in CreateTableOne(vars = vars, data = hints4cycle2, factorVars = factorVars): Dropping variable(s) Age due to unsupported class.

##   
## Overall   
## n 3630   
## GenderC (%)   
## -9 68 ( 1.9)   
## 1 1390 (38.3)   
## 2 2172 (59.8)   
## Education (%)   
## -9 89 ( 2.5)   
## 1 94 ( 2.6)   
## 2 235 ( 6.5)   
## 3 775 (21.3)   
## 4 267 ( 7.4)   
## 5 790 (21.8)   
## 6 845 (23.3)   
## 7 535 (14.7)   
## RaceEthn (%)   
## -9 372 (10.2)   
## 1 511 (14.1)   
## 2 2043 (56.3)   
## 3 496 (13.7)   
## 4 14 ( 0.4)   
## 5 98 ( 2.7)   
## 6 6 ( 0.2)   
## 7 90 ( 2.5)   
## MaritalStatus (%)   
## -9 88 ( 2.4)   
## -5 14 ( 0.4)   
## 1 1712 (47.2)   
## 2 145 ( 4.0)   
## 3 528 (14.5)   
## 4 377 (10.4)   
## 5 138 ( 3.8)   
## 6 628 (17.3)   
## OccupationStatus (%)   
## -9 82 ( 2.3)   
## -5 64 ( 1.8)   
## 1 1808 (49.8)   
## 2 242 ( 6.7)   
## 3 234 ( 6.4)   
## 4 78 ( 2.1)   
## 5 872 (24.0)   
## 6 232 ( 6.4)   
## 91 18 ( 0.5)   
## HHInc (%)   
## -9 480 (13.2)   
## 1 740 (20.4)   
## 2 501 (13.8)   
## 3 459 (12.6)   
## 4 524 (14.4)   
## 5 926 (25.5)   
## MedConditions\_Diabetes (%)   
## -9 133 ( 3.7)   
## -1 11 ( 0.3)   
## 1 659 (18.2)   
## 2 2827 (77.9)   
## MedConditions\_HighBP (%)   
## -9 122 ( 3.4)   
## -1 11 ( 0.3)   
## 1 1499 (41.3)   
## 2 1998 (55.0)   
## MedConditions\_HeartCondition (%)   
## -9 120 ( 3.3)   
## -1 11 ( 0.3)   
## 1 359 ( 9.9)   
## 2 3140 (86.5)   
## SmokeNow (%)   
## -9 26 ( 0.7)   
## -6 26 ( 0.7)   
## -5 1 ( 0.0)   
## -2 299 ( 8.2)   
## -1 1753 (48.3)   
## 1 442 (12.2)   
## 2 144 ( 4.0)   
## 3 939 (25.9)

Here is a messy Table 1, it looks like I will have to go in and recode all of the missing categories to one variable and add labels. And maybe limit this to only those that have had prostate cancer vs those with only a family history vs those with neither.

My research question will be: Is there a difference between fruit and vegetable intake and exercise habits between those with prostate cancer vs those with a family history. This will be a univariate t-test at first. Then I could possibly control for age, smoking status, and education.