Sodium and Blood Pressure

Table of Contents

library(haven)  
library(data.table)  
library(ggplot2)  
  
sav1 <- "UKDA-6533-spss/spss/spss25/ndns\_rp\_yr1-4a\_personleveldietarydata\_uk\_v2.sav"  
sav2 <- "UKDA-6533-spss/spss/spss25/ndns\_rp\_yr1-4a\_indiv\_uk.sav"  
sav3 <- "UKDA-6533-spss/spss/spss25/ndns\_yr1-3indiva\_weights.sav"

#### Does Blood pressure correlate with sodium intake when controlling for income, race and age

In A representative UK population NDNS is raised intake of sodium >rda associated with raised Blood pressure readings?

## Introduction

Sodium is a key electrolyte in cellular physiology. One core function of the cell is to maintain a sodium concentration gradient across the cell wall. Maintaining sodium levels is therefore an essential part of all higher organisms. This role is largely taken by the kidney and modified by receptors and hormones from across the organism. It also relies on ingestion and taste, or diet.

I would like to understand what to tell my patients. Current advice is to reduce sodium intake but it can be difficult to identify the best way to do this. Recent critiques of the role of ingested sodium in blood pressure have looked again at the best form for this advice. The National dietary nutrition dataset is a rolling cross sectional study with linked data on ingestion of sodium and of blood pressure. It also contains data on potential confounding factors such as age sex weight race and income.

Machine learning is a way of interrogating data sets to identify potential models to explain and predict an outcome. In this case the outcome is systolic or diastolic blood pressure.

Causation analysis working with regression models can help to better identify the directional role of particular variables in a model. This helps to identify which variables to include in models and which combination is most likely to give a clinically significant answer.

#savx <- "UKDA-6533-spss/spss/spss25/"  
  
sav1d <- read\_sav(sav1)  
#View(sav1d)  
  
sav2d <- read\_sav(sav2)  
#View(sav2d)  
  
#savxd <- read\_sav(savx)  
   
sav3d <- read\_sav(sav3)  
  
  
  
#yr1-4a person level dietary data\_uk\_v2 data  
persdat <- subset(sav1d[c("seriali","Age", "Sex","Country","Sodiummg","Calciummg","TotalEMJ")])  
sugarset <- subset(sav1d[c("seriali", "Totalsugarsg","Glucoseg","Sucroseg","Fructoseg","Lactoseg","SOFTDRINKSLOWCALORIE","SOFTDRINKSNOTLOWCALORIE", "TEACOFFEEANDWATER")])  
  
persdat <- as.data.table(persdat)  
sugarset <- as.data.table(sugarset)  
  
persdat[, Sex := factor(Sex, levels = 1:2, labels = c("Male", "Female"))]  
persdat[, Country := factor(Country)]  
  
summary(persdat)

## seriali Age Sex Country   
## Min. :10101032 Min. : 1.00 Male :3157 England :3441   
## 1st Qu.:20204104 1st Qu.:10.00 Female:3671 Northern Ireland: 982   
## Median :30208051 Median :20.00 Scotland :1695   
## Mean :27237093 Mean :29.73 Wales : 710   
## 3rd Qu.:40213171 3rd Qu.:48.00   
## Max. :90305261 Max. :96.00   
## Sodiummg Calciummg TotalEMJ   
## Min. : 133.6 Min. : 95.19 Min. : 0.4161   
## 1st Qu.:1528.7 1st Qu.: 583.12 1st Qu.: 5.4673   
## Median :1959.9 Median : 754.97 Median : 6.7228   
## Mean :2067.2 Mean : 794.85 Mean : 7.0154   
## 3rd Qu.:2492.3 3rd Qu.: 961.04 3rd Qu.: 8.2618   
## Max. :7405.5 Max. :3234.23 Max. :22.3664

summary(sugarset)

## seriali Totalsugarsg Glucoseg Sucroseg   
## Min. :10101032 Min. : 3.607 Min. : 0.000 Min. : 0.2567   
## 1st Qu.:20204104 1st Qu.: 65.309 1st Qu.: 9.215 1st Qu.: 25.9719   
## Median :30208051 Median : 87.685 Median :13.975 Median : 38.4110   
## Mean :27237093 Mean : 93.671 Mean :15.838 Mean : 42.5467   
## 3rd Qu.:40213171 3rd Qu.:115.656 3rd Qu.:19.931 3rd Qu.: 54.2214   
## Max. :90305261 Max. :369.092 Max. :86.768 Max. :283.2085   
## Fructoseg Lactoseg SOFTDRINKSLOWCALORIE SOFTDRINKSNOTLOWCALORIE  
## Min. : 0.000 Min. : 0.000 Min. : 0.0 Min. : 0.0   
## 1st Qu.: 8.899 1st Qu.: 6.243 1st Qu.: 0.0 1st Qu.: 0.0   
## Median :13.968 Median :10.436 Median : 0.0 Median : 50.0   
## Mean :15.812 Mean :12.117 Mean : 144.4 Mean : 144.8   
## 3rd Qu.:20.428 3rd Qu.:16.179 3rd Qu.: 198.0 3rd Qu.: 200.0   
## Max. :84.123 Max. :79.406 Max. :5050.8 Max. :2197.5   
## TEACOFFEEANDWATER  
## Min. : 0.0   
## 1st Qu.: 200.0   
## Median : 556.5   
## Mean : 696.1   
## 3rd Qu.:1052.9   
## Max. :5295.9

#measured and recorded data yr1-4a\_indiv\_uk   
#subsets of the table to identify grouped information  
bpset <- subset(sav2d[c("seriali","Sys", "Dias", "Sys2", "Dias2","omsysval","omdiaval","CutIll")])  
ethnset <- subset(sav2d[c("seriali", "EthGrG","EthGrU" ,"ethgr5","ethgr2")])  
saltset <- subset(sav2d[c("seriali","SaltChk","SalHowC","SltSHow", "Na\_mmol", "Na\_mmol\_Corrected" ,"Na\_mmol\_24h\_4\_10CLAIM", "Na\_mmol\_24h\_4\_10CLAIM\_Corrected" )])  
medsset <- subset(sav2d[c("seriali", "bpmedc","bpmedd")])  
hypset <- subset(sav2d[c("seriali", "hyper140", "hibp140", "hyper1", "highbp1")])  
incset <- subset(sav2d[c("seriali","eqvinc","nssec8","hhinc")])  
measset <- subset(sav2d[c("seriali","htval","wtval","bmival")])  
ageset <- subset(sav2d[c("seriali","agegad1", "agegad2", "agegch1","agegr1","age")])

#change subsets to datatables   
 bpset <- as.data.table(bpset)  
 saltset <- as.data.table(saltset)  
 medsset <- as.data.table(medsset)  
 hypset <- as.data.table(hypset)  
 ethnset <- as.data.table(ethnset)  
 incset <- as.data.table(incset)  
 measset <- as.data.table(measset)  
 ageset <- as.data.table(ageset)  
   
 #define factors   
 #saltset  
 saltset[, SaltChk := factor(SaltChk, levels = 1:8, labels = c("Salt",  
 "Salt substitute",  
 "Neither",  
 "Item not applicable",  
 "No answer/refused",  
 "Don't know",  
 "Qn not applicable to survey year",  
 "Schedule not applicable"))]  
 saltset[, SalHowC := factor(SalHowC, levels = 1:8,labels = c("Always",  
"Usually",  
"Sometimes",  
 ",Item not applicable",  
 "No answer/refused",  
 " Don't know",  
"Qn not applicable to survey year",  
"Schedule not applicable"))]  
 saltset[, SltSHow := factor(SltSHow,levels = 1:8,labels = c("Always",  
 "Usually",  
 "Sometimes",  
 ",Item not applicable",  
 "No answer/refused",  
 " Don't know",  
 "Qn not applicable to survey year",  
 "Schedule not applicable"))]  
   
#medset  
 medsset[, bpmedc := factor(bpmedc)]  
 medsset[, bpmedd := factor(bpmedd)]  
   
   
 #hypeset  
 hypset[,hyper140 := factor(hyper140,levels = 1:9, labels = c(" Normotensive untreated",  
" Normotensive treated",  
 "Hypertensive treated",  
"Hypertensive untreated",  
"No answer/refused",  
 "Don't know",  
"Refused, attempted but not obtained, not attempted",  
 "Qn not applicable to survey year",  
"Item not applicable")) ]  
 hypset[, hibp140 := factor(hibp140, levels = 1:7, labels = c("Not high BP",  
"High BP",  
 "No answer/refused",  
 "Don't know",  
 "Refused, attempted but not obtained, not attempted",  
 "Qn not applicable to survey year",  
 "Item not applicable")) ]  
 hypset[, hyper1 := factor(hyper1,levels = 1:9, labels = c(" Normotensive untreated",  
 " Normotensive treated",  
 "Hypertensive treated",  
 "Hypertensive untreated",  
 "No answer/refused",  
 "Don't know",  
 "Refused, attempted but not obtained, not attempted",  
 "Qn not applicable to survey year",  
 "Item not applicable"))]  
 hypset[, highbp1 := factor(highbp1, levels = 1:7, labels = c("Not high BP",  
 "High BP",  
 "No answer/refused",  
 "Don't know",  
 "Refused, attempted but not obtained, not attempted",  
 "Qn not applicable to survey year",  
 "Item not applicable")) ]  
   
   
 cutillcat <- c("Cancer (neoplasm) including lumps, masses, tumours and growths and benign (non-malignant) lumps and cysts"  
 ,"Diabetes. Incl. Hyperglycemia"  
 , "Other endocrine/metabolic"  
 ,"Mental illness/anxiety/depression/nerves (nes)"  
 ,"Mental handicap"  
 , "Epilepsy/fits/convulsions"  
 ,"Migraine/headaches  
 "," Other problems of nervous system  
 "," Cataract/poor eye sight/blindness  
 ","Other eye complaints  
 ","Poor hearing/deafness  
 ","Tinnitus/noises in the ear  
 ","Menieres disease/ear complaints causing balance problems  
"," Other ear complaints  
 "," Stroke/cerebral haemorrhage/cerebral thrombosis  
 ","Heart attack/angina  
","Hypertension/high blood pressure/blood pressure (nes)  
 "," Other heart problems  
 ","Piles/haemorrhoids incl. Varicose Veins in anus  
 "," Varicose veins/phlebitis in lower extremities  
","Other blood vessels/embolic  
 ","Bronchitis/emphysema  
 ","Asthma  
 "," Hayfever  
 ","Other respiratory complaints  
 ","Stomach ulcer/ulcer (nes)/abdominal hernia/rupture  
 ","Other digestive complaints (stomach, liver, pancreas, bile ducts, small intestine)  
 ","Complaints of bowel/colon (large intestine, caecum, bowel, colon, rectum)  
 ","Complaints of teeth/mouth/tongue  
 ","Kidney complaints  
 ","Urinary tract infection  
 ","Other bladder problems/incontinence  
 ","Reproductive system disorders  
 ","Arthritis/rheumatism/fibrositis  
 ","Back problems/slipped disc/spine/neck  
","Other problems of bones/joints/muscles  
"," Infectious and parasitic disease  
","Disorders of blood and blood forming organs and immunity disorders  
 ","Skin complaints  
 ","Other complaints  
","Unclassifiable (no other codable complaint)  
 ","Complaint no longer present  
","No answer/refused  
 ","Dont know"  
 ,"Qn not applicable to survey year"  
 ,"Item not applicable")  
   
#incset  
   
 hhinccat <- c("Under £5,000","  
 £5,000 - £9,999","  
 £10,000 - £14,999","  
 £15,000 - £19,999","  
 £20,000 - £24,999","  
 £25,000 - £29,999","  
 £30,000 - £34,999","  
 £35,000 - £39,999","  
 £40,000 - £44,999","  
 £45,000 - £49,999","  
£50,000 - £74,999","  
 £75,000 - £99,999","  
£100,000 or more","  
 No answer/refused","  
 Dont know","  
Qn not applicable to survey year","  
Item not applicable")  
   
   
#bpset   
 bpset[, CutIll := factor(CutIll, levels = 1:46, labels = cutillcat)]  
 bpset[, "omsysval" := as.numeric(omsysval)]  
 bpset[,Sys := as.numeric(Sys)]  
  
   
#ethnset   
ethnset[ , EthGrG := factor(EthGrG)]  
ethnset[ , EthGrU := factor(EthGrU)]  
ethnset[ , ethgr5 := factor(ethgr5, levels = 1:5, labels = c( 'White'  
 , 'Mixed ethnic group'  
 , 'Black or Black British'  
 , 'Asian or asian British'  
 , 'Any other group'))]  
ethnset[ , ethgr2 := factor(ethgr2, levels = 1:2, labels = c( 'White'  
 , 'Non-white'))]  
  
#incset  
incset[ , nssec8 := factor(nssec8, levels = 1:9, labels = c( "Higher managerial and professional occupations", "Lower managerial and professional occupations"  
 , "Intermediate occupations"  
 , "Small employers and own account workers"  
 , "Lower supervisory and technical occupations"  
 , "Semi-routine occupations"  
 , "Routine occupations"  
 , "Never worked"  
 , "Other"))]  
incset[ , hhinc := factor (hhinc, levels = 1:17,labels = hhinccat)]  
  
  
#ageset  
ageset[,agegad1 := factor(agegad1, levels = 1:4 ,labels = c("16-24","25-49","50-64","65+ years"))]  
ageset[,agegad2 := factor(agegad2, levels = 1:5, labels = c("16-18","19-34","35-49","50-64","65+ years"))]  
ageset[,agegch1 := factor(agegch1, levels = 1:3, labels = c("8-10","11-12","13-15"))]  
ageset[,agegr1 := factor(agegr1, levels = 1:5, labels = c("1.5-3 years","4-10 years","11-18 years","19-64 years","65+ years"))]  
   
#add columns for g equivalent of mmol for sodium  
saltset$Na\_g\_Corrected <- saltset$Na\_mmol\_Corrected/17.1  
saltset$Na\_g <- saltset$Na\_mmol/17.1  
saltset$Na\_g\_24h\_4\_10CLAIM <- saltset$Na\_mmol\_24h\_4\_10CLAIM/17.1  
saltset$Na\_g\_24h\_4\_10CLAIM\_Corrected <- saltset$Na\_mmol\_24h\_4\_10CLAIM\_Corrected/17.1

# the literature

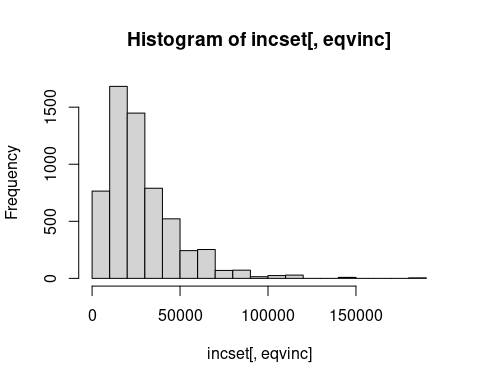
The time trend analysis showed changes in salt intake. Whilst BP measurements were made the results are not reported on in the paper.

It would be useful to understand if the predicted improvements in BP were found in the same population.

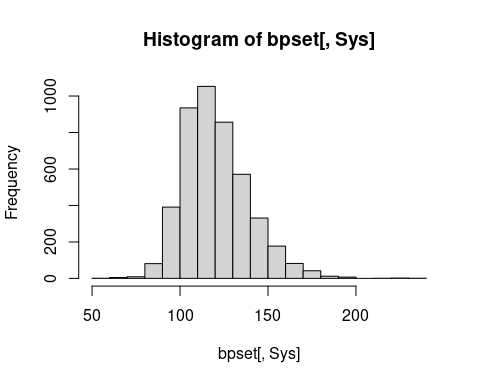
Then the meaning of these results needs to be understandable at the level of those able to implement policy. The NDNS population is structured to match UK age and sex profiles as well as regional representation.

Rebasing this population to match practice, LA or ICB populations will give a better understanding of how the results apply to populations for which commissioners have responsibility.

#View(bpset)  
#description of dataset   
#bpset  
   
hist(incset[,eqvinc])



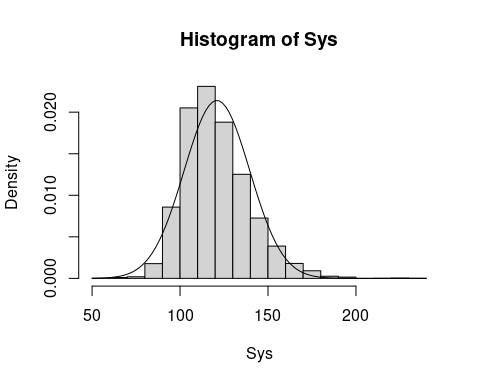
hist(bpset[,Sys])



bpset[, hist(Sys, prob = TRUE)] # histogram

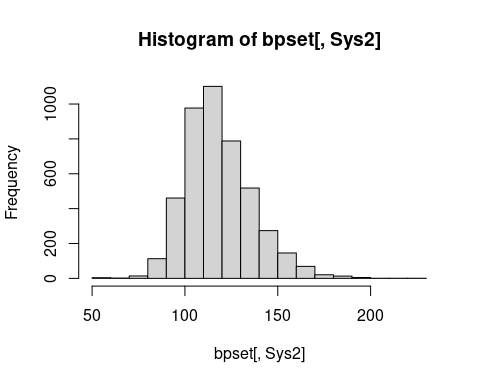
## $breaks  
## [1] 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230  
## [20] 240  
##   
## $counts  
## [1] 1 5 9 81 391 935 1053 857 571 331 177 82 42 12 7  
## [16] 0 1 2 1  
##   
## $density  
## [1] 2.193945e-05 1.096972e-04 1.974550e-04 1.777095e-03 8.578324e-03  
## [6] 2.051338e-02 2.310224e-02 1.880211e-02 1.252742e-02 7.261957e-03  
## [11] 3.883282e-03 1.799035e-03 9.214568e-04 2.632734e-04 1.535761e-04  
## [16] 0.000000e+00 2.193945e-05 4.387889e-05 2.193945e-05  
##   
## $mids  
## [1] 55 65 75 85 95 105 115 125 135 145 155 165 175 185 195 205 215 225 235  
##   
## $xname  
## [1] "Sys"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"

bpset[, curve(  
 dnorm(x, mean(Sys, na.rm = TRUE), sd(Sys, na.rm = TRUE)),  
 add = TRUE)] # superimpose a Normal distribution

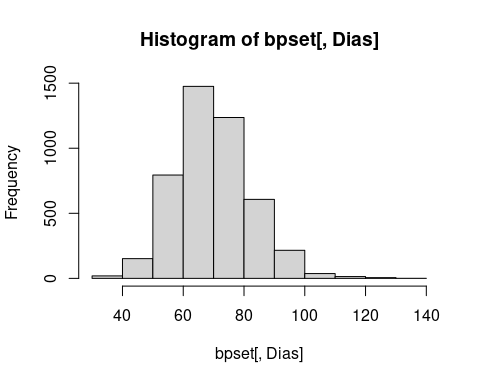


## x y  
## 1: 50.0 1.516881e-05  
## 2: 51.9 2.224979e-05  
## 3: 53.8 3.229877e-05  
## 4: 55.7 4.640147e-05  
## 5: 57.6 6.597254e-05  
## ---   
## 97: 232.4 3.695316e-10  
## 98: 234.3 1.998262e-10  
## 99: 236.2 1.069397e-10  
## 100: 238.1 5.663845e-11  
## 101: 240.0 2.968720e-11

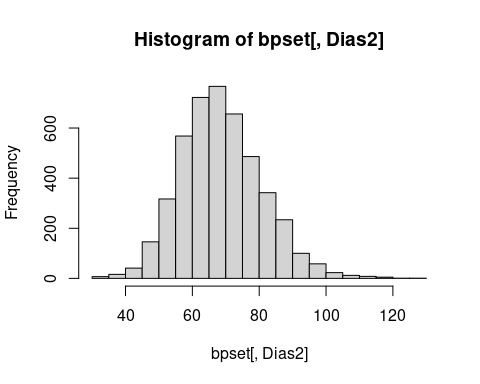
hist(bpset[,Sys2])



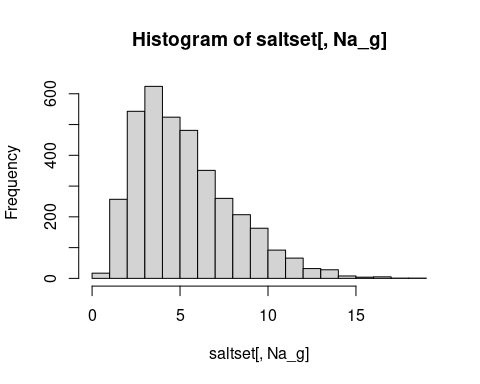
hist(bpset[,Dias])



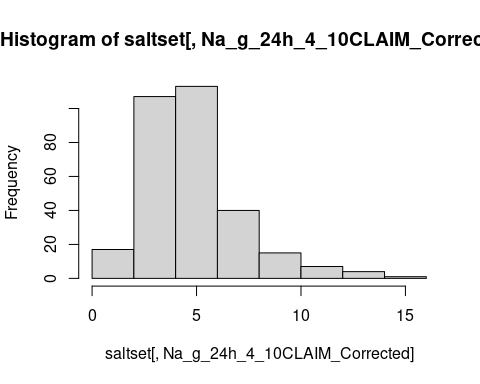
hist(bpset[,Dias2])



hist(saltset[,Na\_g])



hist(saltset[,Na\_g\_24h\_4\_10CLAIM\_Corrected])



#cutillgraph <- ggplot(bpset,aes(CutIll,Sys))+geom\_col()  
#cutillgraph

summary(bpset)

## seriali Sys Dias Sys2   
## Min. :10101032 Min. : 52 Min. : 31.00 Min. : 54   
## 1st Qu.:20204104 1st Qu.:108 1st Qu.: 62.00 1st Qu.:107   
## Median :30208051 Median :118 Median : 69.00 Median :117   
## Mean :27237093 Mean :121 Mean : 70.38 Mean :119   
## 3rd Qu.:40213171 3rd Qu.:132 3rd Qu.: 78.00 3rd Qu.:129   
## Max. :90305261 Max. :234 Max. :137.00 Max. :227   
## NA's :2270 NA's :2271 NA's :2319   
## Dias2 omsysval omdiaval   
## Min. : 31.00 Min. : 77.5 Min. : 35.00   
## 1st Qu.: 61.00 1st Qu.:106.5 1st Qu.: 61.00   
## Median : 69.00 Median :116.0 Median : 68.50   
## Mean : 69.23 Mean :119.1 Mean : 69.24   
## 3rd Qu.: 77.00 3rd Qu.:129.5 3rd Qu.: 76.50   
## Max. :129.00 Max. :220.0 Max. :127.50   
## NA's :2319 NA's :3321 NA's :3321   
## CutIll   
## Other respiratory complaints\n\t : 193   
## Other problems of bones/joints/muscles\n : 127   
## Unclassifiable (no other codable complaint)\n\t : 105   
## Other digestive complaints (stomach, liver, pancreas, bile ducts, small intestine)\n\t: 81   
## Back problems/slipped disc/spine/neck\n : 52   
## (Other) : 245   
## NA's :6025

summary(saltset)

## seriali SaltChk SalHowC   
## Min. :10101032 Salt :3469 Always :1806   
## 1st Qu.:20204104 Neither :3176 Sometimes : 914   
## Median :30208051 Salt substitute : 181 Usually : 747   
## Mean :27237093 Item not applicable: 0 ,Item not applicable: 0   
## 3rd Qu.:40213171 No answer/refused : 0 No answer/refused : 0   
## Max. :90305261 (Other) : 0 (Other) : 0   
## NA's : 2 NA's :3361   
## SltSHow Na\_mmol Na\_mmol\_Corrected  
## Always : 73 Min. : 3.54 Min. : 3.70   
## Usually : 57 1st Qu.: 54.30 1st Qu.: 57.10   
## Sometimes : 51 Median : 80.65 Median : 84.85   
## ,Item not applicable: 0 Mean : 90.03 Mean : 94.71   
## No answer/refused : 0 3rd Qu.:116.92 3rd Qu.:123.03   
## (Other) : 0 Max. :312.20 Max. :328.40   
## NA's :6647 NA's :3164 NA's :3164   
## Na\_mmol\_24h\_4\_10CLAIM Na\_mmol\_24h\_4\_10CLAIM\_Corrected Na\_g\_Corrected   
## Min. : 10.50 Min. : 11.00 Min. : 0.216   
## 1st Qu.: 51.75 1st Qu.: 54.45 1st Qu.: 3.339   
## Median : 71.40 Median : 75.10 Median : 4.962   
## Mean : 77.75 Mean : 81.80 Mean : 5.539   
## 3rd Qu.: 94.72 3rd Qu.: 99.62 3rd Qu.: 7.194   
## Max. :250.30 Max. :263.30 Max. :19.205   
## NA's :6524 NA's :6524 NA's :3164   
## Na\_g Na\_g\_24h\_4\_10CLAIM Na\_g\_24h\_4\_10CLAIM\_Corrected  
## Min. : 0.207 Min. : 0.614 Min. : 0.643   
## 1st Qu.: 3.175 1st Qu.: 3.026 1st Qu.: 3.184   
## Median : 4.716 Median : 4.175 Median : 4.392   
## Mean : 5.265 Mean : 4.547 Mean : 4.783   
## 3rd Qu.: 6.838 3rd Qu.: 5.539 3rd Qu.: 5.826   
## Max. :18.257 Max. :14.637 Max. :15.398   
## NA's :3164 NA's :6524 NA's :6524

summary(medsset)

## seriali bpmedc bpmedd   
## Min. :10101032 0 :1412 0 :1550   
## 1st Qu.:20204104 1 : 450 1 : 312   
## Median :30208051 NA's:4966 NA's:4966   
## Mean :27237093   
## 3rd Qu.:40213171   
## Max. :90305261

summary(hypset)

## seriali hyper140   
## Min. :10101032 Normotensive untreated:2877   
## 1st Qu.:20204104 Hypertensive untreated : 368   
## Median :30208051 Normotensive treated : 142   
## Mean :27237093 Hypertensive treated : 120   
## 3rd Qu.:40213171 No answer/refused : 0   
## Max. :90305261 (Other) : 0   
## NA's :3321   
## hibp140   
## Not high BP : 630   
## High BP : 0   
## No answer/refused : 0   
## Don't know : 0   
## Refused, attempted but not obtained, not attempted: 0   
## (Other) : 0   
## NA's :6198   
## hyper1   
## Normotensive untreated:3149   
## Normotensive treated : 221   
## Hypertensive untreated : 96   
## Hypertensive treated : 41   
## No answer/refused : 0   
## (Other) : 0   
## NA's :3321   
## highbp1   
## Not high BP : 358   
## High BP : 0   
## No answer/refused : 0   
## Don't know : 0   
## Refused, attempted but not obtained, not attempted: 0   
## (Other) : 0   
## NA's :6470

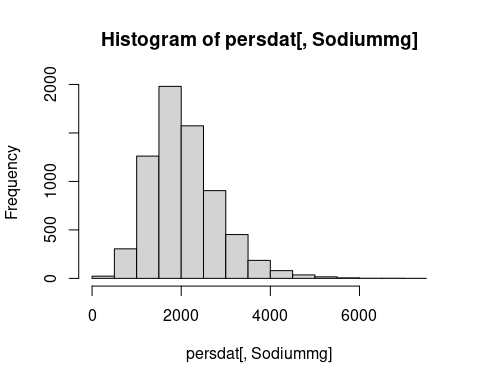
summary(ageset)

## seriali agegad1 agegad2 agegch1   
## Min. :10101032 16-24 : 821 16-18 : 575 8-10 : 546   
## 1st Qu.:20204104 25-49 :1583 19-34 : 770 11-12: 328   
## Median :30208051 50-64 : 868 35-49 :1059 13-15: 594   
## Mean :27237093 65+ years: 753 50-64 : 868 NA's :5360   
## 3rd Qu.:40213171 NA's :2803 65+ years: 753   
## Max. :90305261 NA's :2803   
## agegr1 age   
## 1.5-3 years: 604 Min. : 1.00   
## 4-10 years :1277 1st Qu.:10.00   
## 11-18 years:1497 Median :20.00   
## 19-64 years:2697 Mean :29.73   
## 65+ years : 753 3rd Qu.:48.00   
## Max. :96.00

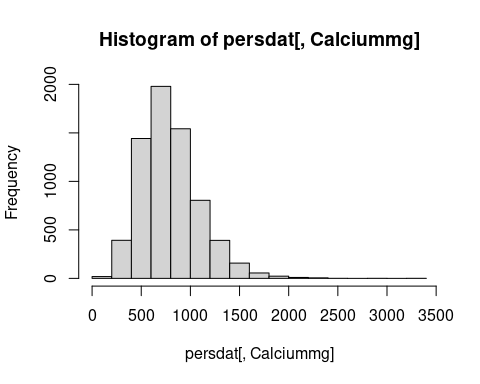
#method The NDNS is a postcode randomised survey which approaches approximately 1000 people each year. The sample selects 500 adults and 500 children. The numbers are managed to deliver a representative sample for the UK.

These participants are asked some basic questions and if they agree to take part are given a 4 day dietary diary. The recorded intake is then analysed for intake by food. These foods have defined contents which can then be reduced to their elemental constituents.

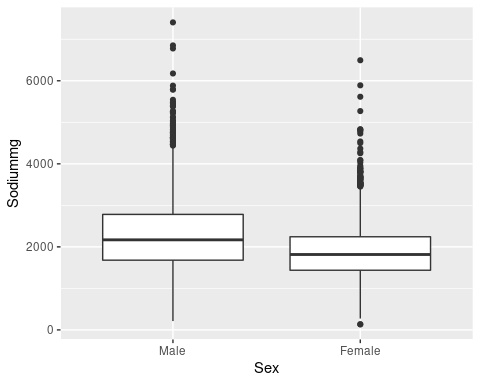
hist(persdat[,Sodiummg])



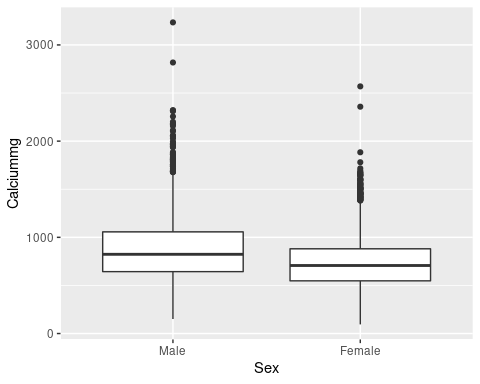
hist(persdat[,Calciummg])



dietnagraph <- ggplot(persdat, aes(Sex,Sodiummg))+ geom\_boxplot()  
dietnagraph



dietcagraph <- ggplot(persdat, aes(Sex,Calciummg))+ geom\_boxplot()  
dietcagraph

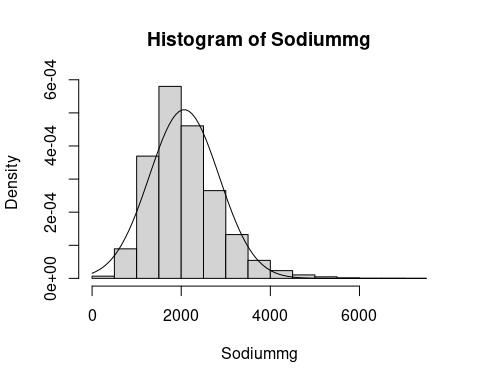


#combine data from tables  
  
persugar <- merge(persdat,sugarset, by = "seriali")  
persethnsugar <- merge(ethnset , persugar, by = "seriali")  
incethsugar <- merge(persethnsugar, incset , by = "seriali")  
  
bpdietdat <- merge(persdat,bpset, by = "seriali" )  
medhypdat <- merge(medsset, hypset, by = "seriali")  
medbppers <- merge(medhypdat, bpset, by ="seriali")  
  
meaage <- merge(measset,ageset, by ="seriali" )  
meaages <- merge(meaage, saltset, by = "seriali")  
  
  
nearlyalldata <- merge(incethsugar, medbppers, by ="seriali")  
alldata <- merge(nearlyalldata, meaages, by ="seriali")  
  
alldatam <- alldata[Sex == "Male"]  
alldataf <- alldata[Sex == "Female"]  
#View(alldata)

#bpdietdatq <- subset(bpdietdat[,c("seriali","Age","Sodiummg","eqvinc","Sys", "Dias", "Sys2", "Dias2","omsysval","omdiaval")])  
#bpdietdatq <- bpdietdatq[,is.na(bpdietdat[Sys]) ]  
#View(bpdietdatq)  
#Summary(bpdietdatq)  
  
bpdietdat[, hist(Sodiummg, prob = TRUE)] # histogram

## $breaks  
## [1] 0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000 5500 6000 6500 7000  
## [16] 7500  
##   
## $counts  
## [1] 23 305 1261 1980 1573 905 452 186 80 36 16 6 2 2 1  
##   
## $density  
## [1] 6.736965e-06 8.933802e-05 3.693615e-04 5.799649e-04 4.607499e-04  
## [6] 2.650849e-04 1.323960e-04 5.448155e-05 2.343292e-05 1.054482e-05  
## [11] 4.686585e-06 1.757469e-06 5.858231e-07 5.858231e-07 2.929115e-07  
##   
## $mids  
## [1] 250 750 1250 1750 2250 2750 3250 3750 4250 4750 5250 5750 6250 6750 7250  
##   
## $xname  
## [1] "Sodiummg"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"

bpdietdat[, curve(  
 dnorm(x, mean(Sodiummg, na.rm = TRUE), sd(Sodiummg, na.rm = TRUE)),  
 add = TRUE  
)] # superimpose a Normal distribution

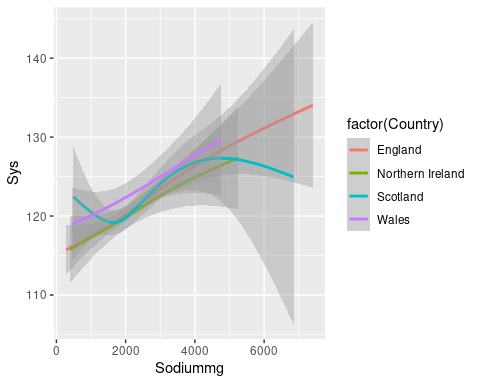


## x y  
## 1: 0 1.556061e-05  
## 2: 75 1.995210e-05  
## 3: 150 2.534902e-05  
## 4: 225 3.191130e-05  
## 5: 300 3.980509e-05  
## ---   
## 97: 7200 2.319412e-13  
## 98: 7275 1.231322e-13  
## 99: 7350 6.477031e-14  
## 100: 7425 3.375913e-14  
## 101: 7500 1.743481e-14

# graph section view the data  
#bpdietdat[Sex == "2" ,plot("Sodiummg","Sys")]  
graph1 <- ggplot(bpdietdat, aes(Sodiummg, Sys, colour = factor(Country)) ) + geom\_smooth()  
graph1

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

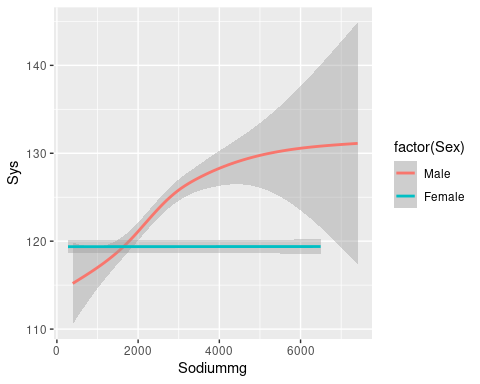
## Warning: Removed 2270 rows containing non-finite values (stat\_smooth).



graph1a <- ggplot(bpdietdat, aes(Sodiummg, Sys, colour = factor(Sex)) )+ geom\_smooth()  
graph1a

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

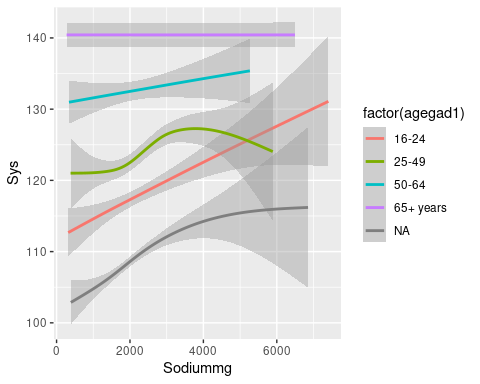
## Warning: Removed 2270 rows containing non-finite values (stat\_smooth).



graph1b <- ggplot(alldata, aes(Sodiummg, Sys, colour = factor(agegad1)) ) + geom\_smooth()  
graph1b

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

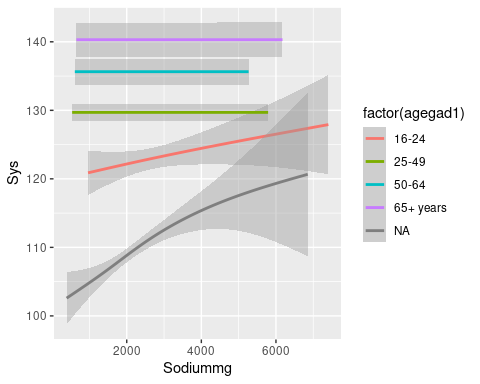
## Warning: Removed 2270 rows containing non-finite values (stat\_smooth).



graph1bm <- ggplot(alldatam, aes(Sodiummg, Sys, colour = factor(agegad1)) ) + geom\_smooth()  
graph1bm

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

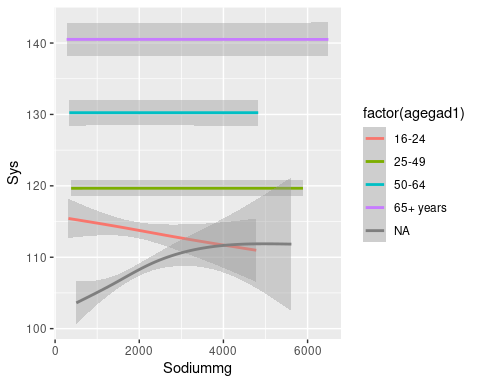
## Warning: Removed 1085 rows containing non-finite values (stat\_smooth).



graph1bf <- ggplot(alldataf, aes(Sodiummg, Sys, colour = factor(agegad1)) ) + geom\_smooth()  
graph1bf

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

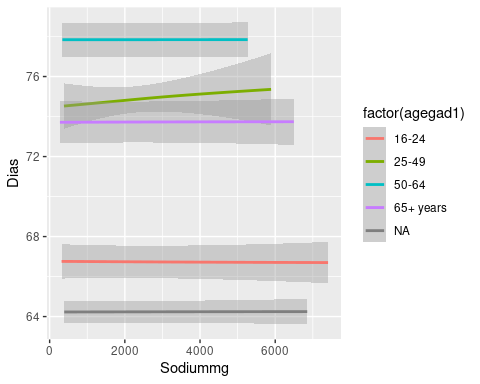
## Warning: Removed 1185 rows containing non-finite values (stat\_smooth).



graph1bd <- ggplot(alldata, aes(Sodiummg, Dias, colour = factor(agegad1)) ) + geom\_smooth()  
graph1bd

## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defaulting to continuous.  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

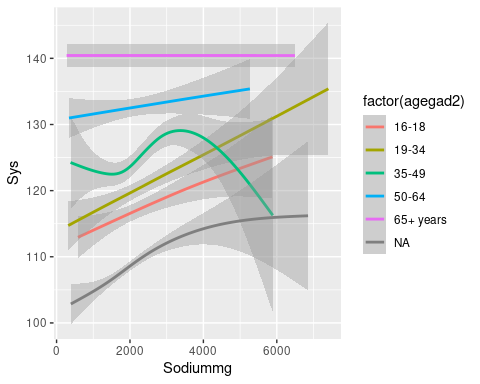
## Warning: Removed 2271 rows containing non-finite values (stat\_smooth).



graph1c <- ggplot(alldata, aes(Sodiummg, Sys, colour = factor(agegad2)) ) + geom\_smooth()  
graph1c

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

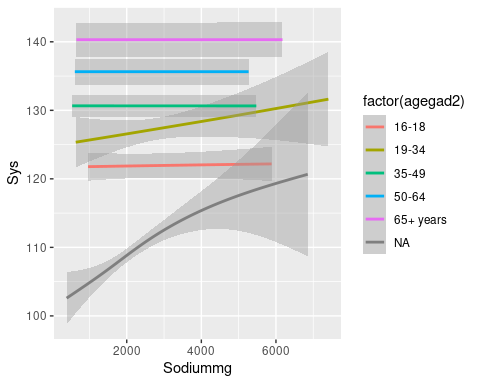
## Warning: Removed 2270 rows containing non-finite values (stat\_smooth).



graph1bm <- ggplot(alldatam, aes(Sodiummg, Sys, colour = factor(agegad2)) ) + geom\_smooth()  
graph1bm

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

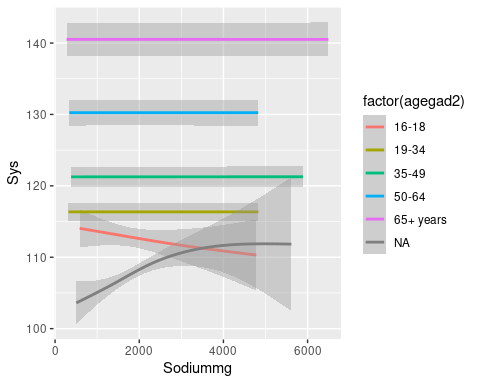
## Warning: Removed 1085 rows containing non-finite values (stat\_smooth).



graph1bf <- ggplot(alldataf, aes(Sodiummg, Sys, colour = factor(agegad2)) ) + geom\_smooth()  
graph1bf

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

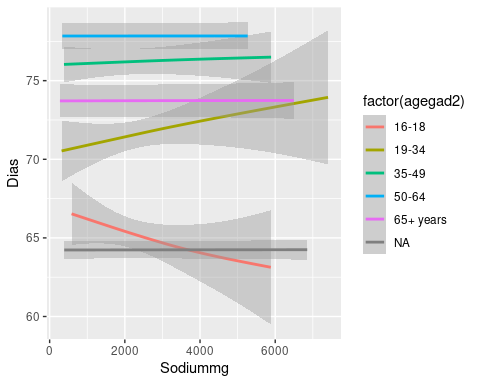
## Warning: Removed 1185 rows containing non-finite values (stat\_smooth).



graph1cd <- ggplot(alldata, aes(Sodiummg, Dias, colour = factor(agegad2)) ) + geom\_smooth()  
graph1cd

## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defaulting to continuous.  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

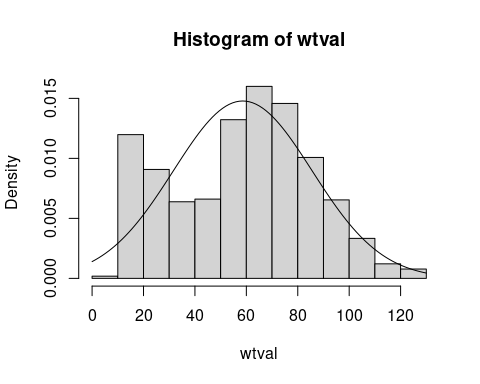
## Warning: Removed 2271 rows containing non-finite values (stat\_smooth).



#bpdietdatq <- subset(bpdietdat[,c("seriali","Age","Sodiummg","eqvinc","Sys", "Dias", "Sys2", "Dias2","omsysval","omdiaval")])  
#bpdietdatq <- bpdietdatq[,is.na(bpdietdat[Sys]) ]  
#View(bpdietdatq)  
#Summary(bpdietdatq)  
  
alldata[, hist(wtval, prob = TRUE)] # histogram

## $breaks  
## [1] 0 10 20 30 40 50 60 70 80 90 100 110 120 130  
##   
## $counts  
## [1] 12 769 583 410 424 849 1027 936 647 420 214 78 50  
##   
## $density  
## [1] 0.0001869450 0.0119800592 0.0090824116 0.0063872877 0.0066053902  
## [6] 0.0132263592 0.0159993768 0.0145817105 0.0100794516 0.0065430752  
## [11] 0.0033338526 0.0012151425 0.0007789375  
##   
## $mids  
## [1] 5 15 25 35 45 55 65 75 85 95 105 115 125  
##   
## $xname  
## [1] "wtval"  
##   
## $equidist  
## [1] TRUE  
##   
## attr(,"class")  
## [1] "histogram"

alldata[, curve(  
 dnorm(x, mean(wtval, na.rm = TRUE), sd(wtval, na.rm = TRUE)),  
 add = TRUE  
)] # superimpose a Normal distribution



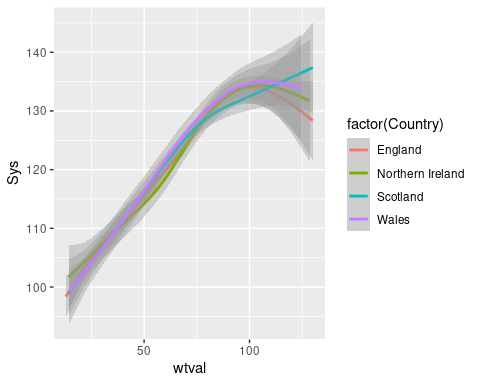
## x y  
## 1: -8.881784e-16 0.0013992842  
## 2: 1.300000e+00 0.0015517911  
## 3: 2.600000e+00 0.0017169302  
## 4: 3.900000e+00 0.0018952395  
## 5: 5.200000e+00 0.0020872171  
## ---   
## 97: 1.248000e+02 0.0007290833  
## 98: 1.261000e+02 0.0006470583  
## 99: 1.274000e+02 0.0005729302  
## 100: 1.287000e+02 0.0005061184  
## 101: 1.300000e+02 0.0004460613

# graph section view the data  
#bpdietdat[Sex == "2" ,plot("Sodiummg","Sys")]  
graphw1 <- ggplot(alldata, aes(wtval, Sys, colour = factor(Country)) ) + geom\_smooth()  
graphw1

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

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

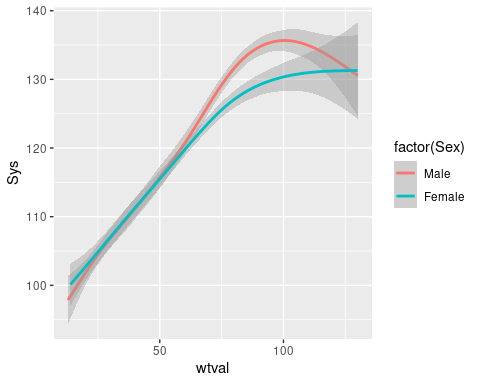
## Warning: Removed 2423 rows containing non-finite values (stat\_smooth).



graphw1a <- ggplot(alldata, aes(wtval, Sys, colour = factor(Sex)) )+ geom\_smooth()  
graphw1a

## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defaulting to continuous.  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

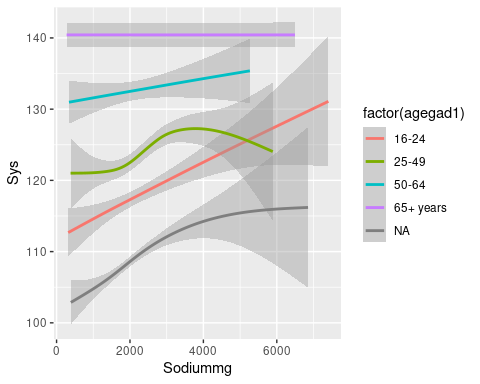
## Warning: Removed 2423 rows containing non-finite values (stat\_smooth).



graph1wb <- ggplot(alldata, aes(wtval, Sys, colour = factor(agegad1)) ) + geom\_smooth()  
graph1b

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

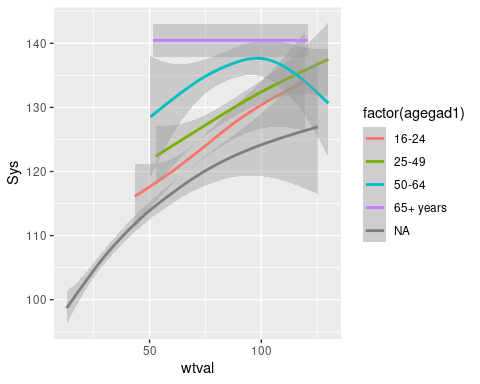
## Warning: Removed 2270 rows containing non-finite values (stat\_smooth).



graphw1bm <- ggplot(alldatam, aes(wtval, Sys, colour = factor(agegad1)) ) + geom\_smooth()  
graphw1bm

## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defaulting to continuous.  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

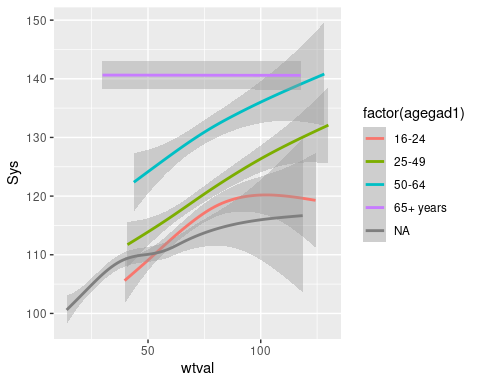
## Warning: Removed 1159 rows containing non-finite values (stat\_smooth).



graphw1bf <- ggplot(alldataf, aes(wtval, Sys, colour = factor(agegad1)) ) + geom\_smooth()  
graphw1bf

## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defaulting to continuous.  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

## Warning: Removed 1264 rows containing non-finite values (stat\_smooth).



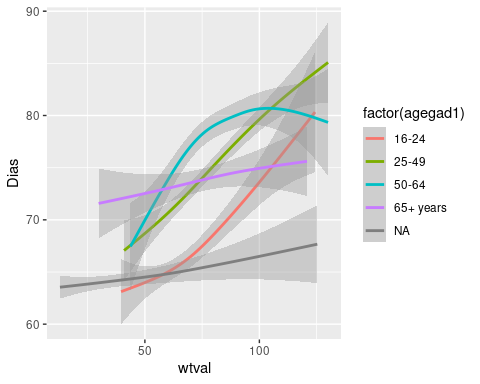
graphw1bd <- ggplot(alldata, aes(wtval, Dias, colour = factor(agegad1)) ) + geom\_smooth()  
graphw1bd

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

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

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

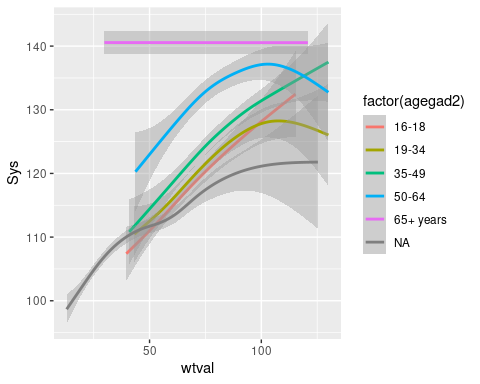
## Warning: Removed 2424 rows containing non-finite values (stat\_smooth).



graphw1c <- ggplot(alldata, aes(wtval, Sys, colour = factor(agegad2)) ) + geom\_smooth()  
graphw1c

## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defaulting to continuous.  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

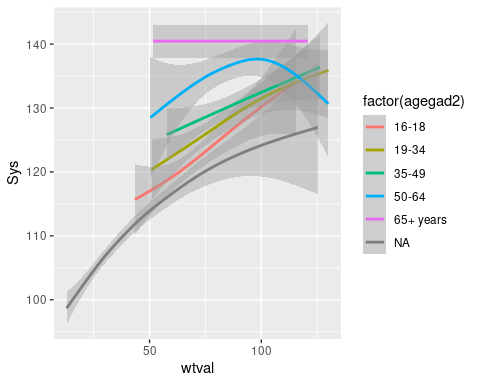
## Warning: Removed 2423 rows containing non-finite values (stat\_smooth).



graphw1bm <- ggplot(alldatam, aes(wtval, Sys, colour = factor(agegad2)) ) + geom\_smooth()  
graphw1bm

## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defaulting to continuous.  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

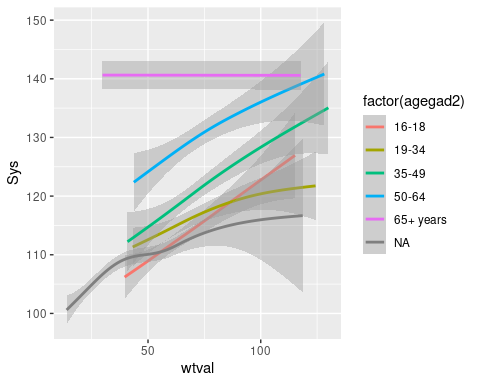
## Warning: Removed 1159 rows containing non-finite values (stat\_smooth).



graphw1bf <- ggplot(alldataf, aes(wtval, Sys, colour = factor(agegad2)) ) + geom\_smooth()  
graphw1bf

## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defaulting to continuous.  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

## Warning: Removed 1264 rows containing non-finite values (stat\_smooth).



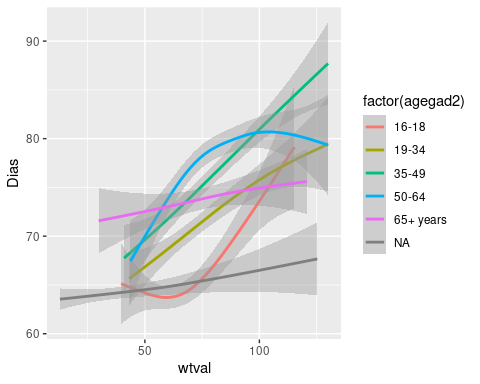
graphw1cd <- ggplot(alldata, aes(wtval, Dias, colour = factor(agegad2)) ) + geom\_smooth()  
graphw1cd

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

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

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

## Warning: Removed 2424 rows containing non-finite values (stat\_smooth).



#discussion

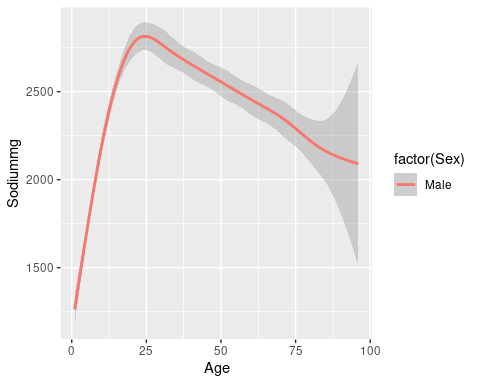
The dataset shows that sodium intake is related to systolic blood pressure in particualr age groups of men. There is much less of a relationship in women.

These relationships are different in different age groups.

The data support consideration of a more complex approach to preventing blood pressure. The simplest message for avoidance is to loose weight. Changing age and Sex are more complex. Salt reduction is an issue for men between 16 and 50, but even here it is more important to maintain a healthy weight than to reduce sodium intake. This may be due to the bodies ability to auto regulate sodium to much higher levels than oral intake allows.

graph2 <- ggplot(alldatam, aes(Age, Sodiummg, colour = factor(Sex)) )+ geom\_smooth()  
graph2

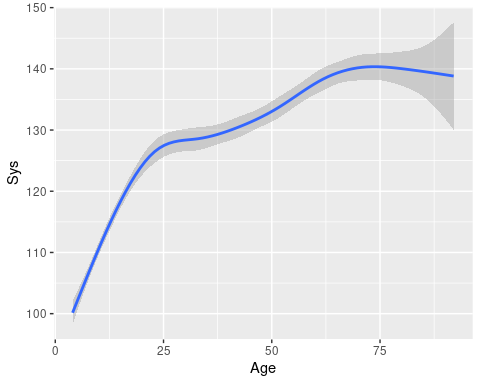
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'



graph2a <- ggplot(alldatam, aes(Age, Sys) ) + geom\_smooth()  
graph2a

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

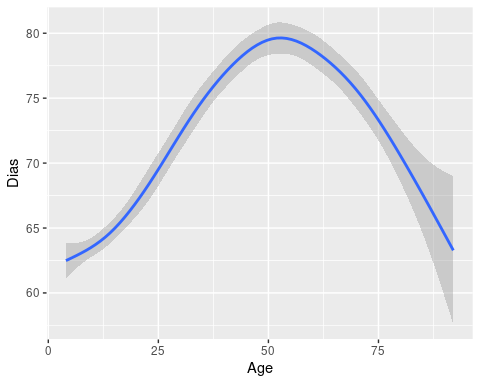
## Warning: Removed 1085 rows containing non-finite values (stat\_smooth).



graph2b <- ggplot(alldatam, aes(Age, Dias) ) + geom\_smooth()  
graph2b

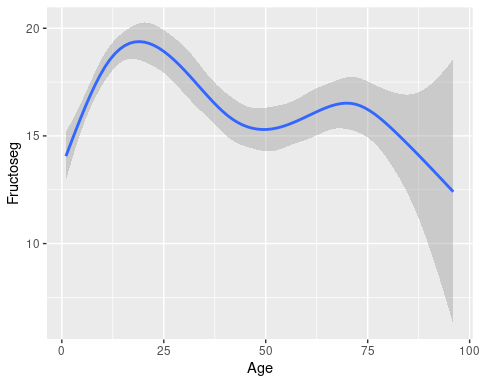
## Don't know how to automatically pick scale for object of type haven\_labelled/vctrs\_vctr/double. Defaulting to continuous.  
## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

## Warning: Removed 1086 rows containing non-finite values (stat\_smooth).

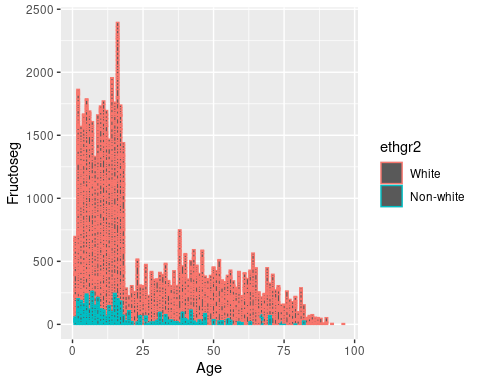


#graph3 <- ggplot(alldatam, aes(Sodiummg, eqvinc) )+ geom\_smooth()  
#graph3  
  
#graph4 <- ggplot(alldatam, aes(Sys, eqvinc) ) +geom\_smooth()  
#graph4  
  
  
  
#show persugstat  
  
graph5 <- ggplot(alldatam, aes(Age, Fructoseg)) + geom\_smooth()  
graph5

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'



#show ethnsugar  
  
graph6 <- ggplot(alldatam, aes (Age , Fructoseg, colour = ethgr2))+geom\_col()  
graph6



# statistical analysis section   
  
# comparison tables  
  
incethtbl <- alldata[,table(nssec8, ethgr2)]  
incethtbl

## ethgr2  
## nssec8 White Non-white  
## Higher managerial and professional occupations 869 79  
## Lower managerial and professional occupations 1602 126  
## Intermediate occupations 576 39  
## Small employers and own account workers 679 58  
## Lower supervisory and technical occupations 648 44  
## Semi-routine occupations 860 94  
## Routine occupations 776 52  
## Never worked 154 32  
## Other 0 0

incethtbl2 <- alldata[,table(nssec8, ethgr5)]  
incethtbl2

## ethgr5  
## nssec8 White Mixed ethnic group  
## Higher managerial and professional occupations 869 17  
## Lower managerial and professional occupations 1602 25  
## Intermediate occupations 576 12  
## Small employers and own account workers 679 11  
## Lower supervisory and technical occupations 648 3  
## Semi-routine occupations 860 24  
## Routine occupations 776 1  
## Never worked 154 5  
## Other 0 0  
## ethgr5  
## nssec8 Black or Black British  
## Higher managerial and professional occupations 12  
## Lower managerial and professional occupations 36  
## Intermediate occupations 10  
## Small employers and own account workers 8  
## Lower supervisory and technical occupations 11  
## Semi-routine occupations 33  
## Routine occupations 18  
## Never worked 9  
## Other 0  
## ethgr5  
## nssec8 Asian or asian British  
## Higher managerial and professional occupations 39  
## Lower managerial and professional occupations 42  
## Intermediate occupations 12  
## Small employers and own account workers 35  
## Lower supervisory and technical occupations 17  
## Semi-routine occupations 27  
## Routine occupations 25  
## Never worked 13  
## Other 0  
## ethgr5  
## nssec8 Any other group  
## Higher managerial and professional occupations 11  
## Lower managerial and professional occupations 23  
## Intermediate occupations 5  
## Small employers and own account workers 4  
## Lower supervisory and technical occupations 13  
## Semi-routine occupations 10  
## Routine occupations 8  
## Never worked 5  
## Other 0

agetbl1 <-alldata[,table(agegr1,Sex)]  
agetbl1

## Sex  
## agegr1 Male Female  
## 1.5-3 years 305 299  
## 4-10 years 665 612  
## 11-18 years 744 753  
## 19-64 years 1126 1571  
## 65+ years 317 436

#how many are on bp medication ?   
medstbl1 <- alldata[ , table(bpmedc,bpmedd)]  
medstbl1

## bpmedd  
## bpmedc 0 1  
## 0 1412 0  
## 1 138 312

hyptabl1 <- alldata[,table(hibp140, bpmedc)]  
hyptabl1

## bpmedc  
## hibp140 0 1  
## Not high BP 180 292  
## High BP 0 0  
## No answer/refused 0 0  
## Don't know 0 0  
## Refused, attempted but not obtained, not attempted 0 0  
## Qn not applicable to survey year 0 0  
## Item not applicable 0 0

hyptabl2 <- alldata[,table(hyper1, bpmedc)]  
hyptabl2

## bpmedc  
## hyper1 0 1  
## Normotensive untreated 970 96  
## Normotensive treated 0 221  
## Hypertensive treated 0 41  
## Hypertensive untreated 48 12  
## No answer/refused 0 0  
## Don't know 0 0  
## Refused, attempted but not obtained, not attempted 0 0  
## Qn not applicable to survey year 0 0  
## Item not applicable 0 0

hyptabl3 <- alldata[,table(hyper140, bpmedc)]  
hyptabl3

## bpmedc  
## hyper140 0 1  
## Normotensive untreated 838 78  
## Normotensive treated 0 142  
## Hypertensive treated 0 120  
## Hypertensive untreated 180 30  
## No answer/refused 0 0  
## Don't know 0 0  
## Refused, attempted but not obtained, not attempted 0 0  
## Qn not applicable to survey year 0 0  
## Item not applicable 0 0

hyptabl4 <- alldata[,table(highbp1, bpmedc)]  
hyptabl4

## bpmedc  
## highbp1 0 1  
## Not high BP 48 274  
## High BP 0 0  
## No answer/refused 0 0  
## Don't know 0 0  
## Refused, attempted but not obtained, not attempted 0 0  
## Qn not applicable to survey year 0 0  
## Item not applicable 0 0

#random correlations  
  
meds <- t.test(medstbl1 , na.rm = TRUE)  
  
incSys <- t.test( hyptabl1, na.rm = TRUE)  
incSys

##   
## One Sample t-test  
##   
## data: hyptabl1  
## t = 1.4259, df = 13, p-value = 0.1775  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## -17.36693 84.79550  
## sample estimates:  
## mean of x   
## 33.71429

meds

##   
## One Sample t-test  
##   
## data: medstbl1  
## t = 1.4461, df = 3, p-value = 0.2439  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## -558.903 1489.903  
## sample estimates:  
## mean of x   
## 465.5

#linear regression models  
lm1 <- lm(Sys ~ agegad2 + Sex +Sodiummg + Calciummg + Fructoseg + eqvinc + ethgr2 +TotalEMJ +wtval + bpmedd, alldata)  
summary(lm1)

##   
## Call:  
## lm(formula = Sys ~ agegad2 + Sex + Sodiummg + Calciummg + Fructoseg +   
## eqvinc + ethgr2 + TotalEMJ + wtval + bpmedd, data = alldata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -51.188 -10.878 -1.425 9.314 91.544   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.164e+02 3.596e+00 32.368 < 2e-16 \*\*\*  
## agegad219-34 2.419e+00 2.156e+00 1.122 0.261976   
## agegad235-49 8.248e+00 2.042e+00 4.038 5.72e-05 \*\*\*  
## agegad250-64 1.377e+01 2.030e+00 6.783 1.83e-11 \*\*\*  
## agegad265+ years 2.182e+01 2.034e+00 10.731 < 2e-16 \*\*\*  
## SexFemale -4.973e+00 1.180e+00 -4.216 2.67e-05 \*\*\*  
## Sodiummg 6.433e-05 9.305e-04 0.069 0.944897   
## Calciummg -3.855e-03 2.089e-03 -1.846 0.065200 .   
## Fructoseg -5.880e-02 4.849e-02 -1.213 0.225467   
## eqvinc 1.023e-05 2.336e-05 0.438 0.661347   
## ethgr2Non-white -2.083e+00 2.420e+00 -0.861 0.389549   
## TotalEMJ 1.588e-01 3.783e-01 0.420 0.674782   
## wtval 8.578e-02 3.094e-02 2.772 0.005658 \*\*   
## bpmedd1 4.434e+00 1.337e+00 3.316 0.000941 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 16.43 on 1212 degrees of freedom  
## (5602 observations deleted due to missingness)  
## Multiple R-squared: 0.244, Adjusted R-squared: 0.2359   
## F-statistic: 30.08 on 13 and 1212 DF, p-value: < 2.2e-16

lmD1 <- lm(Dias ~ agegad2 + Sex +Sodiummg + Calciummg + Fructoseg + eqvinc + ethgr2 +TotalEMJ +wtval + bpmedd, alldata)  
summary(lmD1)

##   
## Call:  
## lm(formula = Dias ~ agegad2 + Sex + Sodiummg + Calciummg + Fructoseg +   
## eqvinc + ethgr2 + TotalEMJ + wtval + bpmedd, data = alldata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -30.338 -7.251 -0.417 7.033 53.604   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.460e+01 2.376e+00 22.983 < 2e-16 \*\*\*  
## agegad219-34 3.660e+00 1.424e+00 2.569 0.01030 \*   
## agegad235-49 8.992e+00 1.350e+00 6.663 4.07e-11 \*\*\*  
## agegad250-64 1.003e+01 1.341e+00 7.477 1.45e-13 \*\*\*  
## agegad265+ years 6.525e+00 1.344e+00 4.855 1.36e-06 \*\*\*  
## SexFemale 1.420e+00 7.795e-01 1.822 0.06874 .   
## Sodiummg 2.374e-04 6.149e-04 0.386 0.69951   
## Calciummg -4.239e-03 1.380e-03 -3.071 0.00218 \*\*   
## Fructoseg -6.314e-02 3.204e-02 -1.971 0.04900 \*   
## eqvinc 4.585e-06 1.543e-05 0.297 0.76645   
## ethgr2Non-white 3.575e+00 1.599e+00 2.235 0.02557 \*   
## TotalEMJ 4.727e-01 2.500e-01 1.891 0.05885 .   
## wtval 1.585e-01 2.045e-02 7.751 1.92e-14 \*\*\*  
## bpmedd1 -2.793e-01 8.837e-01 -0.316 0.75201   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.86 on 1212 degrees of freedom  
## (5602 observations deleted due to missingness)  
## Multiple R-squared: 0.1537, Adjusted R-squared: 0.1446   
## F-statistic: 16.93 on 13 and 1212 DF, p-value: < 2.2e-16

lm2 <- lm(Sys ~ Na\_g\_Corrected + Sex + age + TEACOFFEEANDWATER, alldata)  
summary(lm2)

##   
## Call:  
## lm(formula = Sys ~ Na\_g\_Corrected + Sex + age + TEACOFFEEANDWATER,   
## data = alldata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -57.943 -9.099 -1.066 8.482 83.734   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.058e+02 8.557e-01 123.599 < 2e-16 \*\*\*  
## Na\_g\_Corrected 2.877e-01 9.185e-02 3.133 0.00175 \*\*   
## SexFemale -4.877e+00 4.870e-01 -10.014 < 2e-16 \*\*\*  
## age 4.903e-01 1.284e-02 38.185 < 2e-16 \*\*\*  
## TEACOFFEEANDWATER 2.444e-04 4.800e-04 0.509 0.61073   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 14.48 on 3611 degrees of freedom  
## (3212 observations deleted due to missingness)  
## Multiple R-squared: 0.3776, Adjusted R-squared: 0.3769   
## F-statistic: 547.6 on 4 and 3611 DF, p-value: < 2.2e-16

lmD2 <- lm(Dias ~ Na\_g\_Corrected + Sex + age + TEACOFFEEANDWATER, alldata)  
summary(lmD2)

##   
## Call:  
## lm(formula = Dias ~ Na\_g\_Corrected + Sex + age + TEACOFFEEANDWATER,   
## data = alldata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -39.333 -7.283 -0.579 6.598 65.101   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 62.153383 0.666838 93.206 < 2e-16 \*\*\*  
## Na\_g\_Corrected 0.198544 0.071557 2.775 0.00555 \*\*   
## SexFemale -0.002458 0.379429 -0.006 0.99483   
## age 0.188483 0.010004 18.840 < 2e-16 \*\*\*  
## TEACOFFEEANDWATER 0.001426 0.000374 3.813 0.00014 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.28 on 3610 degrees of freedom  
## (3213 observations deleted due to missingness)  
## Multiple R-squared: 0.1488, Adjusted R-squared: 0.1478   
## F-statistic: 157.7 on 4 and 3610 DF, p-value: < 2.2e-16

library(caret)

## Loading required package: lattice

set.seed(42)  
  
alldataSys <- alldata[!is.na(alldata$Sys),]  
alldataSys <- as.data.frame(alldataSys)  
simplealldata <- subset(alldataSys[c("Sys", "Age","wtval","Sex")])  
  
   
modelSys <- train(Sys ~ ., simplealldata,  
 method = "lm",  
 trControl = trainControl(method = "cv",  
 number = 10,  
 verboseIter = TRUE), na.action = na.pass)

## + Fold01: intercept=TRUE   
## - Fold01: intercept=TRUE   
## + Fold02: intercept=TRUE   
## - Fold02: intercept=TRUE   
## + Fold03: intercept=TRUE   
## - Fold03: intercept=TRUE   
## + Fold04: intercept=TRUE   
## - Fold04: intercept=TRUE   
## + Fold05: intercept=TRUE   
## - Fold05: intercept=TRUE   
## + Fold06: intercept=TRUE   
## - Fold06: intercept=TRUE   
## + Fold07: intercept=TRUE   
## - Fold07: intercept=TRUE   
## + Fold08: intercept=TRUE   
## - Fold08: intercept=TRUE   
## + Fold09: intercept=TRUE   
## - Fold09: intercept=TRUE   
## + Fold10: intercept=TRUE   
## - Fold10: intercept=TRUE   
## Aggregating results  
## Fitting final model on full training set

modelSys

## Linear Regression   
##   
## 4558 samples  
## 3 predictor  
##   
## No pre-processing  
## Resampling: Cross-Validated (10 fold)   
## Summary of sample sizes: 4103, 4102, 4103, 4101, 4102, 4102, ...   
## Resampling results:  
##   
## RMSE Rsquared MAE   
## 14.16787 0.415836 10.70645  
##   
## Tuning parameter 'intercept' was held constant at a value of TRUE

p <- predict(modelSys, simplealldata)  
error <- p - simplealldata$Sys

## Warning in p - simplealldata$Sys: longer object length is not a multiple of  
## shorter object length

rmse\_xval <- sqrt(mean(error^2)) ## xval RMSE  
rmse\_xval

## [1] 22.0307