EDA

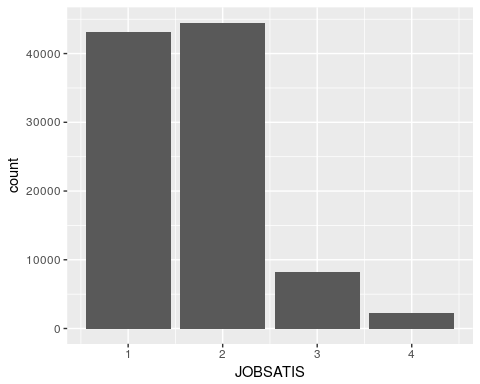
Job satisfaction

ed\_data %>%  
 count(JOBSATIS)

## # A tibble: 4 x 2  
## JOBSATIS n  
## <int+lbl> <int>  
## 1 1 43147  
## 2 2 44508  
## 3 3 8178  
## 4 4 2218

ggplot(data = ed\_data) +   
 geom\_bar(mapping = aes(x = JOBSATIS))

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



##### 

Demographics #####

Age

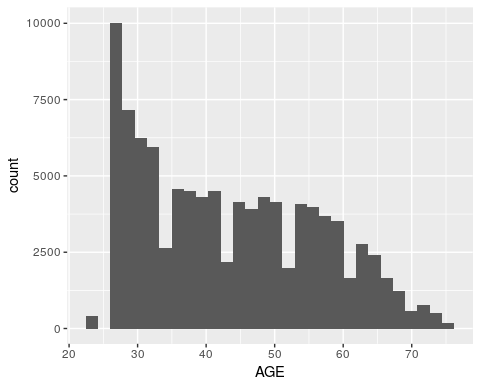
favstats(~AGE, data=ed\_data)

## min Q1 median Q3 max mean sd n missing  
## 23 32 42 54 75 43.44698 12.79204 98051 0

ggplot(data=ed\_data) +   
 geom\_histogram(mapping = aes(x=AGE))

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

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

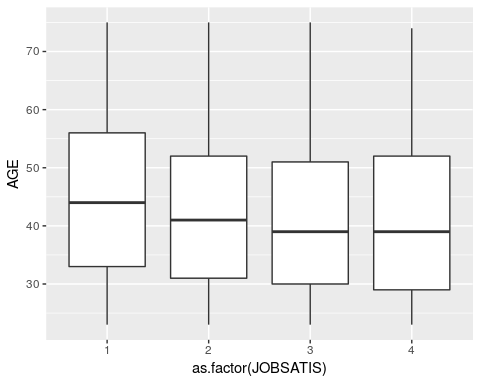


favstats(AGE~JOBSATIS, data=ed\_data)

## JOBSATIS min Q1 median Q3 max mean sd n missing  
## 1 1 23 33 44 56 75 44.97159 13.34282 43147 0  
## 2 2 23 31 41 52 75 42.48697 12.24612 44508 0  
## 3 3 23 30 39 51 75 41.27342 11.87885 8178 0  
## 4 4 23 29 39 52 74 41.06718 12.42531 2218 0

ggplot(data = ed\_data) +   
 geom\_boxplot(mapping = aes(x = as.factor(JOBSATIS), y=AGE))

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



Gender

#Gender  
prop.table(table(ed\_data$GENDER))

##   
## Female Male   
## 0.4242996 0.5757004

prop.table(table(ed\_data$GENDER, ed\_data$JOBSATIS),1)

##   
## 1 2 3 4  
## Female 0.42532990 0.46030334 0.08847920 0.02588756  
## Male 0.45089286 0.44922761 0.07966624 0.02021329

prop.table(table(ed\_data$GENDER, ed\_data$JOBSATIS),2)

##   
## 1 2 3 4  
## Female 0.4101096 0.4302597 0.4501101 0.4855726  
## Male 0.5898904 0.5697403 0.5498899 0.5144274

gender.prop <- prop.table(table(ed\_data$JOBSATIS, ed\_data$GENDER))  
mosaicplot(gender.prop, color=TRUE,main="Job satisfaction by gender")



Race

prop.table(table(ed\_data$RACETH))

##   
## Asian White   
## 0.1739911 0.6188208   
## Under-represented minorities   
## 0.2071881

prop.table(table(ed\_data$RACETH, ed\_data$JOBSATIS),1)

##   
## 1 2 3 4  
## Asian 0.34355217 0.54402110 0.09185229 0.02057444  
## White 0.47850880 0.42560485 0.07632342 0.01956292  
## Under-represented minorities 0.40620231 0.46285996 0.09746493 0.03347280

prop.table(table(ed\_data$RACETH, ed\_data$JOBSATIS),2)

##   
## 1 2 3 4  
## Asian 0.1358379 0.2085243 0.1916116 0.1582507  
## White 0.6729089 0.5802103 0.5662754 0.5351668  
## Under-represented minorities 0.1912532 0.2112654 0.2421130 0.3065825

race.prop <- prop.table(table(ed\_data$JOBSATIS, ed\_data$RACETH))  
mosaicplot(race.prop, color=TRUE,main="Job satisfaction by race")



##### 

Education #####

Highest degree

prop.table(table(ed\_data$DGRDG))

##   
## Bachelor's Master's Doctorate Professional   
## 0.3731120 0.2929292 0.2964988 0.0374601

prop.table(table(ed\_data$DGRDG, ed\_data$JOBSATIS),1)

##   
## 1 2 3 4  
## Bachelor's 0.40736387 0.47419637 0.09044938 0.02799038  
## Master's 0.43604206 0.46034399 0.08296776 0.02064619  
## Doctorate 0.47165658 0.43316593 0.07725647 0.01792102  
## Professional 0.54669208 0.36618568 0.06534168 0.02178056

prop.table(table(ed\_data$DGRDG, ed\_data$JOBSATIS),2)

##   
## 1 2 3 4  
## Bachelor's 0.34540061 0.38977263 0.40462216 0.46167719  
## Master's 0.29026352 0.29707019 0.29139154 0.26735798  
## Doctorate 0.31779730 0.28293790 0.27463928 0.23489630  
## Professional 0.04653858 0.03021929 0.02934703 0.03606853

degree.prop <- prop.table(table(ed\_data$JOBSATIS, ed\_data$DGRDG))  
mosaicplot(degree.prop, color=TRUE,main="Job satisfaction by highest degree")

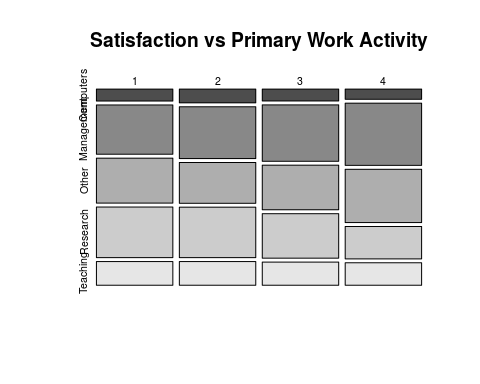


Satisfaction verses Primary Work Activity

ed\_data <- ed\_data %>% mutate(WAPRSM\_non = ifelse(WAPRSM ==1, "Research",  
 ifelse(WAPRSM == 2, "Teaching",  
 ifelse(WAPRSM == 3,"Management",   
 ifelse(WAPRSM ==4, "Computers", "Other")))))  
  
  
waprsm <- prop.table(table(ed\_data$JOBSATIS, ed\_data$WAPRSM\_non),1)  
waprsm

##   
## Computers Management Other Research Teaching  
## 1 0.06616914 0.27292743 0.24910191 0.28039029 0.13141122  
## 2 0.07614361 0.28702705 0.22604925 0.27952278 0.13125730  
## 3 0.06566398 0.31254585 0.24688188 0.24651504 0.12839325  
## 4 0.05635708 0.34445446 0.29531109 0.17944094 0.12443643

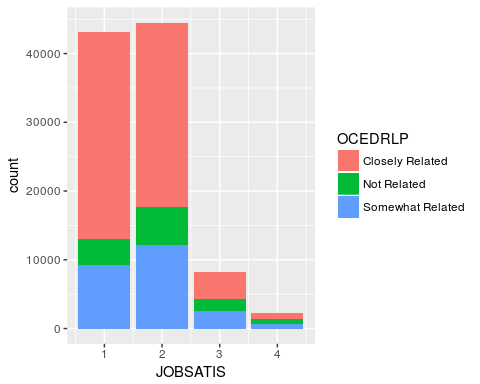
mosaicplot(waprsm, color=TRUE,main="Satisfaction vs Primary Work Activity")



Job Related To Major

g <- ggplot(ed\_data, aes(JOBSATIS, fill= OCEDRLP)) + geom\_bar()  
g

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



plyr::count(ed\_data, "OCEDRLP")

## OCEDRLP freq  
## 1 Closely Related 61602  
## 2 Not Related 12052  
## 3 Somewhat Related 24397

table(ed\_data$JOBSATIS, ed\_data$OCEDRLP)

##   
## Closely Related Not Related Somewhat Related  
## 1 30103 3874 9170  
## 2 26764 5597 12147  
## 3 3944 1744 2490  
## 4 791 837 590

Primary Job Work

table(ed\_data$NOCPRMG\_non, ed\_data$JOBSATIS)

##   
## 1 2 3 4  
## Bio 3616 3819 717 171  
## Engineers 6385 7097 1096 184  
## MSCS 4698 5615 906 212  
## NS-Engineer-related 13157 13297 2965 1083  
## Physics 2412 2532 455 84  
## S-Engineer-related 9214 8944 1485 360  
## Social 3665 3204 554 124

prop.table(table(ed\_data$NOCPRMG, ed\_data$JOBSATIS), 2) #Divides by Column

##   
## 1 2  
## Biological, agricultural and other life scientists 0.08380652 0.08580480  
## Computer and mathematical scientists 0.10888358 0.12615710  
## Engineers 0.14798248 0.15945448  
## Non-science and engineering occupations 0.30493429 0.29875528  
## Physical and related scientists 0.05590192 0.05688865  
## Science and engineering related occupations 0.21354903 0.20095264  
## Social and related scientists 0.08494217 0.07198706  
##   
## 3 4  
## Biological, agricultural and other life scientists 0.08767425 0.07709648  
## Computer and mathematical scientists 0.11078503 0.09558161  
## Engineers 0.13401810 0.08295762  
## Non-science and engineering occupations 0.36255808 0.48827773  
## Physical and related scientists 0.05563708 0.03787196  
## Science and engineering related occupations 0.18158474 0.16230839  
## Social and related scientists 0.06774272 0.05590622

prop.table(table(ed\_data$NOCPRMG, ed\_data$JOBSATIS),1) #Divides by Row

##   
## 1 2  
## Biological, agricultural and other life scientists 0.43445873 0.45884897  
## Computer and mathematical scientists 0.41098767 0.49120812  
## Engineers 0.43252947 0.48076141  
## Non-science and engineering occupations 0.43134876 0.43593863  
## Physical and related scientists 0.43990516 0.46179099  
## Science and engineering related occupations 0.46063091 0.44713293  
## Social and related scientists 0.48562343 0.42453955  
##   
## 3 4  
## Biological, agricultural and other life scientists 0.08614682 0.02054548  
## Computer and mathematical scientists 0.07925816 0.01854606  
## Engineers 0.07424468 0.01246444  
## Non-science and engineering occupations 0.09720674 0.03550587  
## Physical and related scientists 0.08298377 0.01532008  
## Science and engineering related occupations 0.07423886 0.01799730  
## Social and related scientists 0.07340665 0.01643037

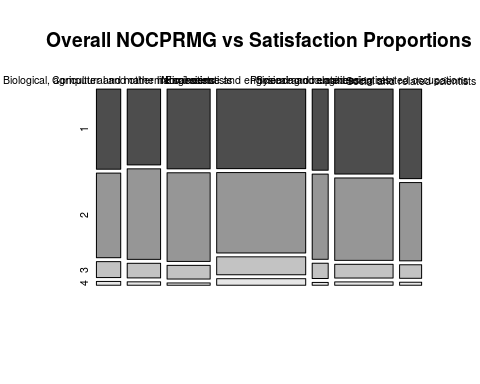
prop.table(table(ed\_data$NOCPRMG, ed\_data$JOBSATIS)) #Overall

##   
## 1  
## Biological, agricultural and other life scientists 0.036878767  
## Computer and mathematical scientists 0.047913841  
## Engineers 0.065119173  
## Non-science and engineering occupations 0.134185271  
## Physical and related scientists 0.024599443  
## Science and engineering related occupations 0.093971505  
## Social and related scientists 0.037378507  
##   
## 2  
## Biological, agricultural and other life scientists 0.038949118  
## Computer and mathematical scientists 0.057266117  
## Engineers 0.072380700  
## Non-science and engineering occupations 0.135613099  
## Physical and related scientists 0.025823296  
## Science and engineering related occupations 0.091217836  
## Social and related scientists 0.032676872  
##   
## 3  
## Biological, agricultural and other life scientists 0.007312521  
## Computer and mathematical scientists 0.009240089  
## Engineers 0.011177856  
## Non-science and engineering occupations 0.030239365  
## Physical and related scientists 0.004640442  
## Science and engineering related occupations 0.015145180  
## Social and related scientists 0.005650121  
##   
## 4  
## Biological, agricultural and other life scientists 0.001743990  
## Computer and mathematical scientists 0.002162140  
## Engineers 0.001876574  
## Non-science and engineering occupations 0.011045272  
## Physical and related scientists 0.000856697  
## Science and engineering related occupations 0.003671559  
## Social and related scientists 0.001264648

prop.table(table(ed\_data$NOCPRMG, ed\_data$JOBSATIS)) #Divides by Column

##   
## 1  
## Biological, agricultural and other life scientists 0.036878767  
## Computer and mathematical scientists 0.047913841  
## Engineers 0.065119173  
## Non-science and engineering occupations 0.134185271  
## Physical and related scientists 0.024599443  
## Science and engineering related occupations 0.093971505  
## Social and related scientists 0.037378507  
##   
## 2  
## Biological, agricultural and other life scientists 0.038949118  
## Computer and mathematical scientists 0.057266117  
## Engineers 0.072380700  
## Non-science and engineering occupations 0.135613099  
## Physical and related scientists 0.025823296  
## Science and engineering related occupations 0.091217836  
## Social and related scientists 0.032676872  
##   
## 3  
## Biological, agricultural and other life scientists 0.007312521  
## Computer and mathematical scientists 0.009240089  
## Engineers 0.011177856  
## Non-science and engineering occupations 0.030239365  
## Physical and related scientists 0.004640442  
## Science and engineering related occupations 0.015145180  
## Social and related scientists 0.005650121  
##   
## 4  
## Biological, agricultural and other life scientists 0.001743990  
## Computer and mathematical scientists 0.002162140  
## Engineers 0.001876574  
## Non-science and engineering occupations 0.011045272  
## Physical and related scientists 0.000856697  
## Science and engineering related occupations 0.003671559  
## Social and related scientists 0.001264648

mosaicplot(prop.table(table(ed\_data$NOCPRMG, ed\_data$JOBSATIS)), color=TRUE,main="Overall NOCPRMG vs Satisfaction Proportions")



Interaction Between Relationship to Major and Primary Job Work

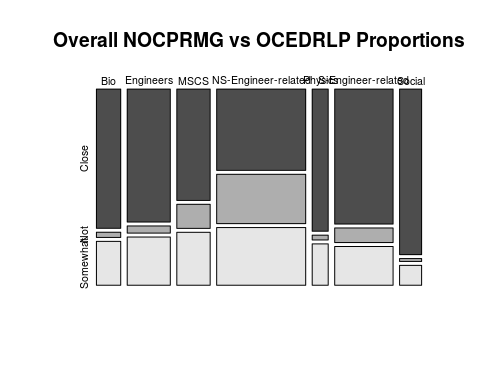
ed\_data <- ed\_data %>% mutate(OCEDRLP\_non = ifelse(OCEDRLP =="Closely Related", "Close",  
 ifelse(OCEDRLP == "Somewhat Related", "Somewhat", "Not")))  
  
prop.table(table(ed\_data$NOCPRMG\_non, ed\_data$OCEDRLP\_non),2) #Divides by Column

##   
## Close Not Somewhat  
## Bio 0.099883121 0.019166943 0.079476985  
## Engineers 0.169150352 0.046714238 0.154896094  
## MSCS 0.109687997 0.121141719 0.131737509  
## NS-Engineer-related 0.213678127 0.663126452 0.383120876  
## Physics 0.067189377 0.011533355 0.049391319  
## S-Engineer-related 0.232719717 0.128609359 0.168750256  
## Social 0.107691309 0.009707932 0.032626962

prop.table(table(ed\_data$NOCPRMG\_non, ed\_data$OCEDRLP\_non),1)

##   
## Close Not Somewhat  
## Bio 0.73927670 0.02775442 0.23296888  
## Engineers 0.70586641 0.03813846 0.25599512  
## MSCS 0.59111189 0.12772286 0.28116525  
## NS-Engineer-related 0.43154547 0.26201561 0.30643892  
## Physics 0.75487872 0.02535109 0.21977020  
## S-Engineer-related 0.71669250 0.07748838 0.20581913  
## Social 0.87902478 0.01550285 0.10547237

mosaicplot(prop.table(table(ed\_data$NOCPRMG\_non, ed\_data$OCEDRLP\_non)), color=TRUE,main="Overall NOCPRMG vs OCEDRLP Proportions")



#### 

Work Related ####

Hours/week worked

prop.table(table(ed\_data$HRSWKGR))

##   
## 20 or less 21 - 35 36 - 40 Greater than 40   
## 0.07495079 0.08271206 0.37754842 0.46478873

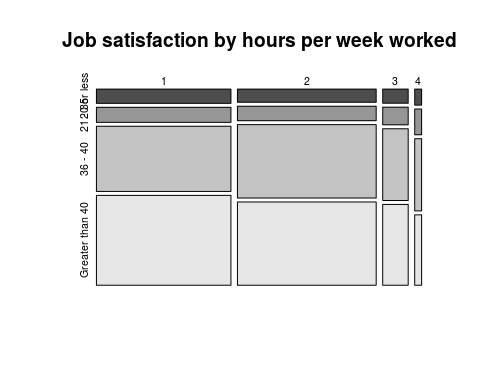
prop.table(table(ed\_data$HRSWKGR, ed\_data$JOBSATIS),1)

##   
## 1 2 3 4  
## 20 or less 0.45489182 0.43352837 0.08531773 0.02626208  
## 21 - 35 0.43378545 0.43156597 0.09654747 0.03810111  
## 36 - 40 0.41189659 0.47864610 0.08598287 0.02347443  
## Greater than 40 0.46163298 0.44111645 0.07866500 0.01858557

prop.table(table(ed\_data$HRSWKGR, ed\_data$JOBSATIS),2)

##   
## 1 2 3 4  
## 20 or less 0.07747931 0.07158264 0.07666911 0.08701533  
## 21 - 35 0.08153522 0.07863755 0.09574468 0.13931470  
## 36 - 40 0.35339653 0.39810821 0.38921497 0.39179441  
## Greater than 40 0.48758894 0.45167161 0.43837124 0.38187556

hour.prop <- prop.table(table(ed\_data$JOBSATIS, ed\_data$HRSWKGR))  
mosaicplot(hour.prop, color=TRUE,main="Job satisfaction by hours per week worked")

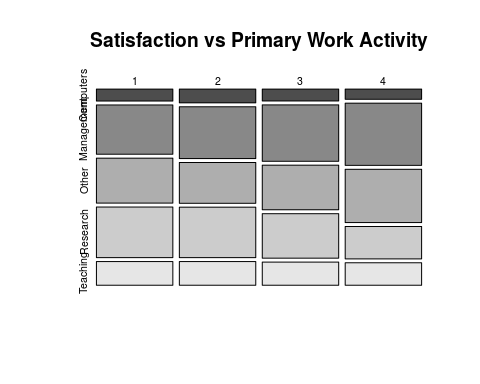


Satisfaction verses Primary Work Activity

ed\_data <- ed\_data %>% mutate(WAPRSM\_non = ifelse(WAPRSM ==1, "Research",  
 ifelse(WAPRSM == 2, "Teaching",  
 ifelse(WAPRSM == 3,"Management",   
 ifelse(WAPRSM ==4, "Computers", "Other")))))  
  
  
waprsm <- prop.table(table(ed\_data$JOBSATIS, ed\_data$WAPRSM\_non),1)  
waprsm

##   
## Computers Management Other Research Teaching  
## 1 0.06616914 0.27292743 0.24910191 0.28039029 0.13141122  
## 2 0.07614361 0.28702705 0.22604925 0.27952278 0.13125730  
## 3 0.06566398 0.31254585 0.24688188 0.24651504 0.12839325  
## 4 0.05635708 0.34445446 0.29531109 0.17944094 0.12443643

mosaicplot(waprsm, color=TRUE,main="Satisfaction vs Primary Work Activity")



Interaction of Available Benefits Compared to Insurnace

library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

ed\_data <- ed\_data %>% mutate(JOBPENS = ifelse(JOBPENS ==1, "Yes", "No"))  
ed\_data <- ed\_data %>% mutate(JOBVAC = ifelse(JOBVAC ==1, "Yes", "No"))  
  
pens.prop <- prop.table(table(ed\_data$JOBINS, ed\_data$JOBPENS))  
vac.prop <- prop.table(table(ed\_data$JOBINS, ed\_data$JOBVAC))  
proft.prop <- prop.table(table(ed\_data$JOBINS, ed\_data$JOBPROFT))

Interaction of Available Benefits Compared to Insurnace

library(gridExtra)   
  
ed\_data <- ed\_data %>% mutate(JOBPENS = ifelse(JOBPENS ==1, "Yes", "No"))  
ed\_data <- ed\_data %>% mutate(JOBVAC = ifelse(JOBVAC ==1, "Yes", "No"))  
  
pens.prop <- prop.table(table(ed\_data$JOBINS, ed\_data$JOBPENS))  
vac.prop <- prop.table(table(ed\_data$JOBINS, ed\_data$JOBVAC))  
proft.prop <- prop.table(table(ed\_data$JOBINS, ed\_data$JOBPROFT))

Proportion of Insurance vs Pension Plan

pens.prop

##   
## No  
## No 0.1529816  
## Yes 0.8470184

Proportion of Insurance vs Paid Vacation/Sick/Leave Days

vac.prop

##   
## No  
## No 0.1529816  
## Yes 0.8470184

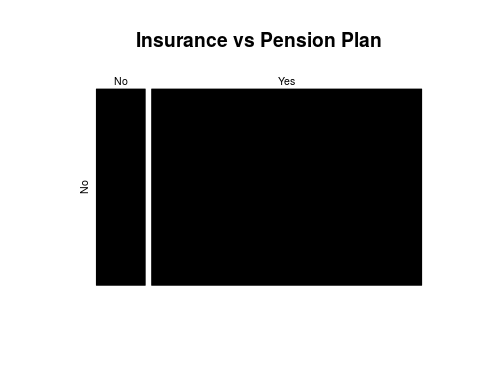
Proportion of Insurance vs Profit Plan

proft.prop

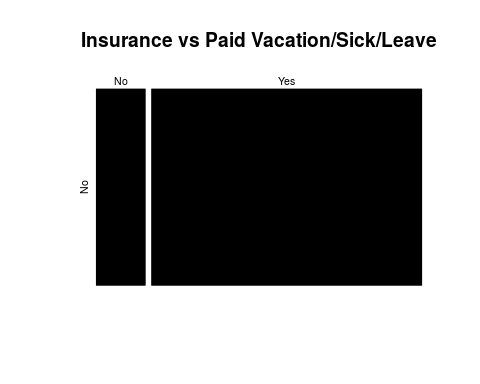
##   
## No Yes  
## No 0.14094706 0.01203455  
## Yes 0.61339507 0.23362332

Mosaic Plots

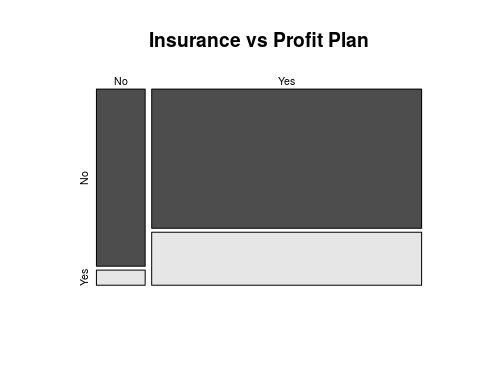
m1<- mosaicplot(pens.prop, color=TRUE,main="Insurance vs Pension Plan")



m2 <- mosaicplot(vac.prop, color=TRUE,main="Insurance vs Paid Vacation/Sick/Leave")



m3 <- mosaicplot(proft.prop, color=TRUE,main="Insurance vs Profit Plan")



Proportion of Job Satisfaction vs Pension Plan

pens1.prop

##   
## No  
## 1 0.44004651  
## 2 0.45392704  
## 3 0.08340557  
## 4 0.02262088

Proportion of Job Satisfaction vs Vacation/Sick/Leave Days

vac1.prop

##   
## No  
## 1 0.44004651  
## 2 0.45392704  
## 3 0.08340557  
## 4 0.02262088

Proportion of Job Satisfaction vs Profit Plan

proft1.prop

##   
## No Yes  
## 1 0.320210911 0.119835596  
## 2 0.347064283 0.106862755  
## 3 0.068015625 0.015389950  
## 4 0.019051310 0.003569571

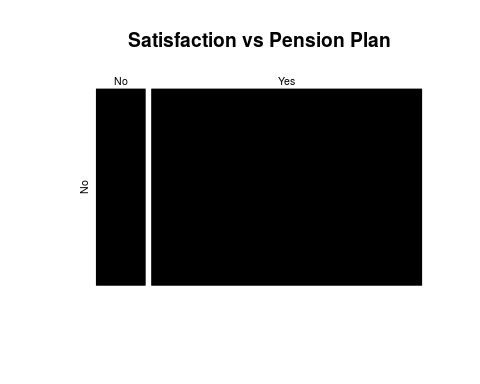
Proportion of Job Satisfaction vs Insurance Plan

ins1.prop

##   
## No Yes  
## 1 0.066801970 0.373244536  
## 2 0.065741298 0.388185740  
## 3 0.014971800 0.068433774  
## 4 0.005466543 0.017154338

Satisfaction verses Available Benefits Mosaic Plots

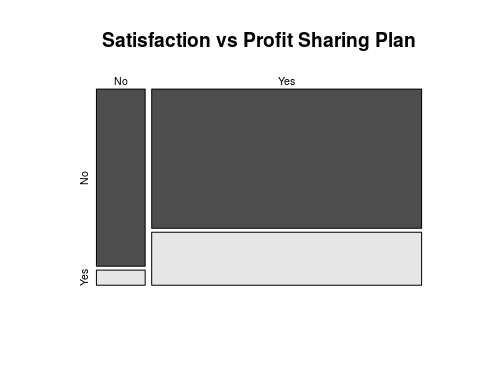
mosaicplot(pens.prop, color=TRUE,main="Satisfaction vs Pension Plan")



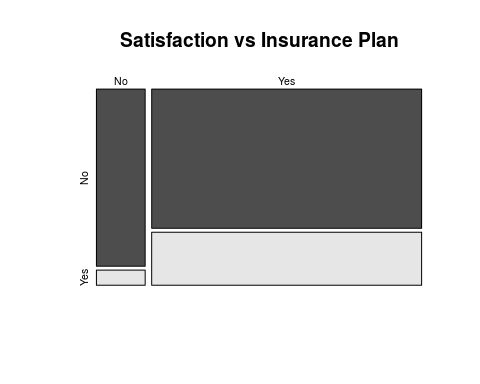
mosaicplot(vac.prop, color=TRUE,main="Satisfaction vs Paid Vacation/Sick/Leave")



mosaicplot(proft.prop, color=TRUE,main="Satisfaction vs Profit Sharing Plan")



mosaicplot(proft.prop, color=TRUE,main="Satisfaction vs Insurance Plan")



Proportion of Insurance vs Pension Plan

pens.prop

##   
## No  
## No 0.1529816  
## Yes 0.8470184

Proportion of Insurance vs Paid Vacation/Sick/Leave Days

vac.prop

##   
## No  
## No 0.1529816  
## Yes 0.8470184

Proportion of Insurance vs Profit Plan

proft.prop

##   
## No Yes  
## No 0.14094706 0.01203455  
## Yes 0.61339507 0.23362332

Salary

library(plyr)

## -------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

## -------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following object is masked from 'package:mosaic':  
##   
## count

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

## The following object is masked from 'package:purrr':  
##   
## compact

ed\_data <- ed\_data %>% mutate(JOBSATIS\_cat = ifelse(JOBSATIS ==1, "Very Satisfied",  
 ifelse(JOBSATIS ==2, "Somewhat Satisfied",  
 ifelse(JOBSATIS ==3, "Somewhat Disatisfied","Very Disatisfied"))))  
  
  
s\_meds <- ddply(ed\_data, .(JOBSATIS), summarise, med = median(SALARY))  
  
ggplot(ed\_data, aes(x=as.factor(JOBSATIS), y=SALARY)) + geom\_boxplot() + xlab("Job Satisfaction") + ylab("Salary") + geom\_text(data = s\_meds, aes(x = JOBSATIS, y = med, label = med), size = 3, vjust = -1.5) + ggtitle("Job Satisfaction vs Salary")

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

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

