college\_mcat<-select(college,major,major\_category,sample\_size,total,median)

m<-0

for(i in 1:nrow(college\_mcat)){

for(j in 1:college\_mcat$total[i]){

m<-m+1

x[m]<-college\_mcat$major\_category[i]

y[m]<-college\_mcat$median[i]

}

}

data<-data.frame(x,y)

pasta\_exp<-lm(y~x,data)

summary(pasta\_exp)

Call:

lm(formula = y ~ x, data = data)

Residuals:

Min 1Q Median 3Q Max

-19696 -6604 -1277 6345 63396

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 41765.68 35.28 1183.97 <2e-16 \*\*\*

xArts -8837.12 39.03 -226.44 <2e-16 \*\*\*

xBiology & Life Science -488.97 38.26 -12.78 <2e-16 \*\*\*

xBusiness 4838.58 36.34 133.14 <2e-16 \*\*\*

xCommunications & Journalism 3914.12 38.70 101.13 <2e-16 \*\*\*

xComputers & Mathematics -3111.00 39.71 -78.33 <2e-16 \*\*\*

xEducation -3803.66 37.71 -100.85 <2e-16 \*\*\*

xEngineering -1558.57 37.81 -41.22 <2e-16 \*\*\*

xHealth 5930.78 38.20 155.25 <2e-16 \*\*\*

xHumanities & Liberal Arts -8569.17 37.20 -230.35 <2e-16 \*\*\*

xIndustrial Arts & Consumer Services -4607.70 40.96 -112.50 <2e-16 \*\*\*

xInterdisciplinary -14265.68 96.64 -147.62 <2e-16 \*\*\*

xLaw & Public Policy -6592.12 42.43 -155.37 <2e-16 \*\*\*

xPhysical Sciences -2382.10 42.20 -56.45 <2e-16 \*\*\*

xPsychology & Social Work -2120.19 38.10 -55.65 <2e-16 \*\*\*

xSocial Science -4389.41 37.84 -115.99 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9976 on 6775999 degrees of freedom

Multiple R-squared: 0.1938, Adjusted R-squared: 0.1938

F-statistic: 1.086e+05 on 15 and 6775999 DF, p-value: < 2.2e-16

nrow(data[x=="xArts"])

[1] 6776015

nrow(data[x=="Arts",])

[1] 357130

sum(college\_mcat[college\_mcat$major\_category=="Arts",]$total)

[1] 357130

nrow(data[x=="Health",])

463230

sum(college\_mcat[college\_mcat$major\_category=="Health",]$total)

463230

data\_avg<-summarize(group\_by(data,x),mean=mean(y))

data\_avg<-arrange(data\_avg,avginc)

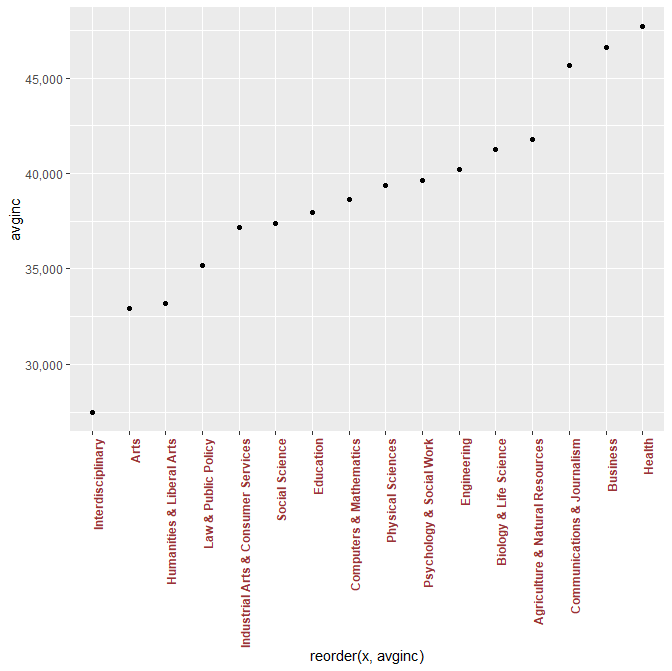
gg<-ggplot(data\_avg,aes(x=reorder(x,avginc),y=avginc))

gg1<-gg+geom\_point()

gg2<-gg1+scale\_y\_continuous(labels = scales::comma)

gg3<-gg2+theme(**axis.text.x = element\_text(face="bold", color="#993333",hjust=1, size=9, angle=90))**

gg3



h<-ggplot(data,aes(x=x,y=y,fill=x))

h1<-h+geom\_point()

h2<-h1+geom\_boxplot()

h3<-h2+theme(axis.text.x = element\_blank())+scale\_y\_continuous(labels = scales::comma)

h3

