

Lab 6

2. `ann<-read.table("agingStudy1FCortexAffyAnn.txt",header=T,row.names=1)`
`affy<-read.table("agingStudy1FCortexAffy.txt",header=T, row.names=1)`
- 3.

```
#Order by age
```

```
ann_age= ann[order(ann$Age),]
```

```
#Paste age into datanames
```

```
age <- paste(dimnames(ann_age)[[1]], ann_age[, 2], ann_age[, 1], sep = ".")
```

```
#New affy dataset
```

```
new_affy= affy[, age]
```

- 4.

```
t.test.all.genes <- function(x,s1,s2) {
```

```
  x1 <- x[s1]
```

```
  x2 <- x[s2]
```

```
  x1 <- as.numeric(x1)
```

```
  x2 <- as.numeric(x2)
```

```
  t.out <- t.test(x1,x2, alternative= c("two.sided"), var.equal=T)
```

```
  out <- as.numeric(t.out$p.value)
```

```
  return(out)
```

```
}
```

```
rawp.gender <- apply(affy[g.g, ],1,t.test.all.genes,s1=ann$Gender == "M",s2=ann$Gender == "F")
```

```
rawp.age <- apply(new_affy[g.a, ],1,t.test.all.genes,s1= ann_age$Age < 50, s2= ann_age$Age > 50)
```

```
  p.corage <- p.adjust(rawp.age,method="holm")
```

```
  p.cor <- p.adjust(rawp.gender,method="holm")
```

- 5.

```
sortageadjusted <- apply(data.frame(p.corage), 2, sort)
```

```
sortgenderadjusted <- apply(data.frame(p.cor), 2, sort)
```

```
sortage <- apply(data.frame(rawp.gender), 2, sort)
```

```
sortgender <- apply(data.frame(rawp.age), 2, sort)
```

```
agevalue <- cbind(sortage, sortageadjusted)
```

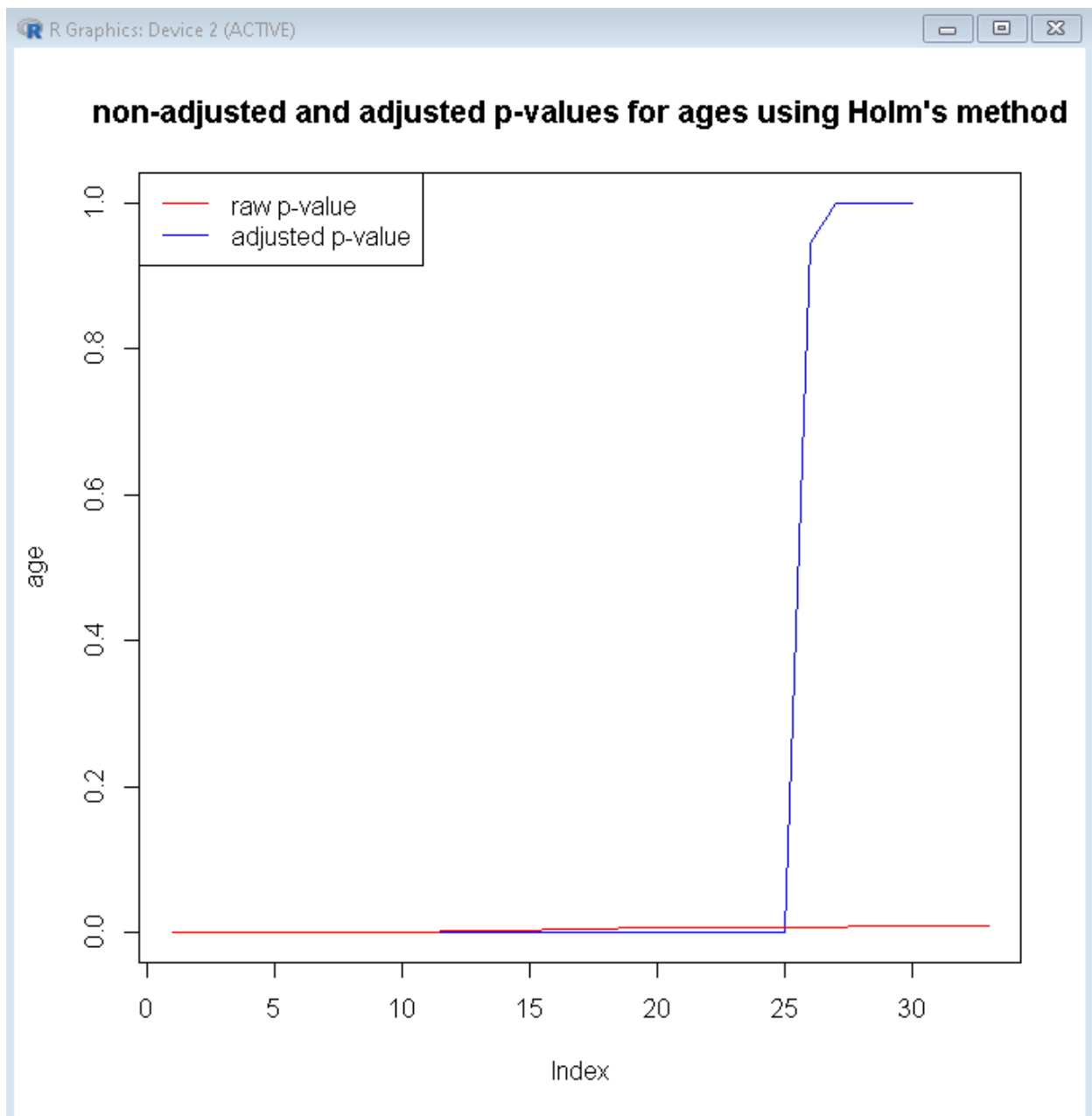
```
  plot(sortage, type = "l", ylim=c(0,1), ylab= "age")
```

```
  lines(sortageadjusted, col="blue")
```

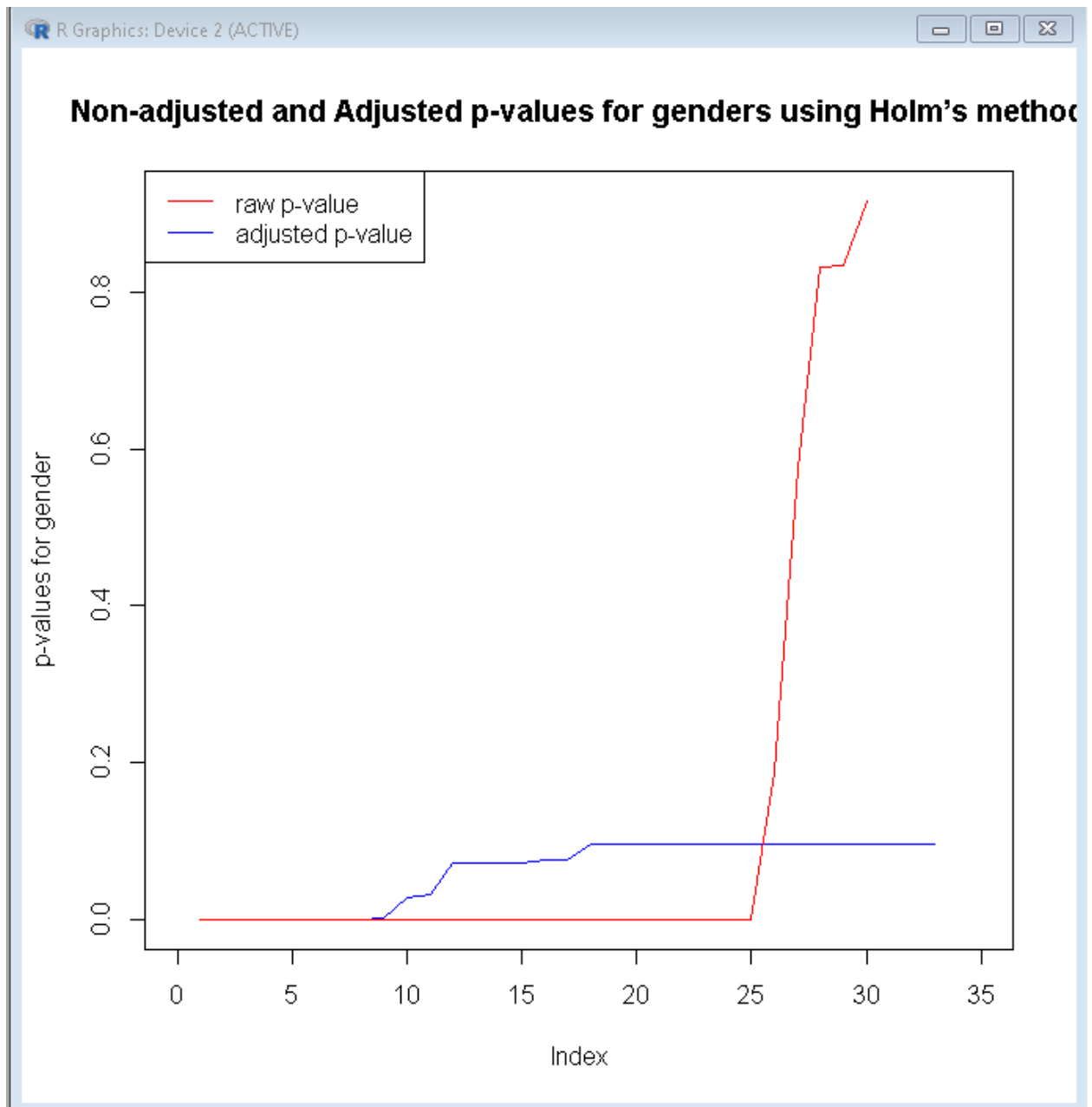
```
  lines(sortage,col="red")
```

```
  title("non-adjusted and adjusted p-values for ages using Holm's method")
```

```
  legend("topleft", c("raw p-value", "adjusted p-value"), lty = c("solid", "solid"), col = c("red", "blue"))
```



```
plot(sortgender, type = "l", ylab = "p-values for gender", xlim=c(0,35))
lines(sortgenderadjusted, col="blue")
title("Non-adjusted and Adjusted p-values for genders using Holm's method")
lines(sortgender, col="red")
legend("topleft", c("raw p-value", "adjusted p-value"), lty = c("solid", "solid"), col = c("red", "blue"))
```



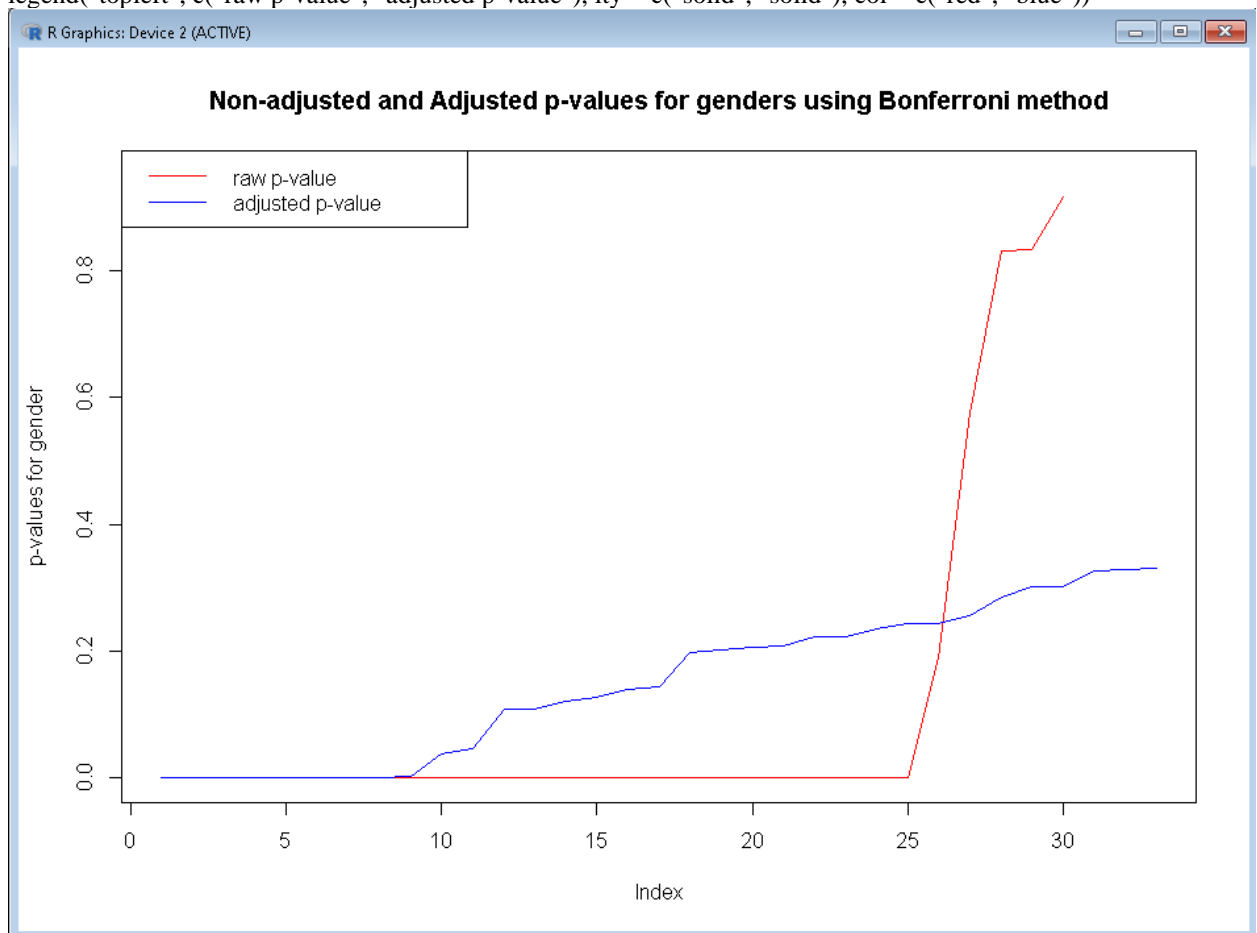
```
6.  
p.corage2 <- p.adjust(rawp.age,method="bonferroni")  
p.cor2<- p.adjust(rawp.gender,method="bonferroni")
```

```
sortagea2 <- apply(data.frame(p.corage2), 2, sort)  
sortgendera2 <- apply(data.frame(p.cor2), 2, sort)  
sortage2 <- apply(data.frame(rawp.gender), 2, sort)  
sortgender2 <- apply(data.frame(rawp.age), 2, sort)
```

```

plot(sortgendera2, type = "l", ylab = "p-values for gender", ylim=c(0,0.95))
lines(sortgender2, col="red")
lines(sortgendera2, col="blue")
title("Non-adjusted and Adjusted p-values for genders using Bonferroni method")
legend("topleft", c("raw p-value", "adjusted p-value"), lty = c("solid", "solid"), col = c("red", "blue"))

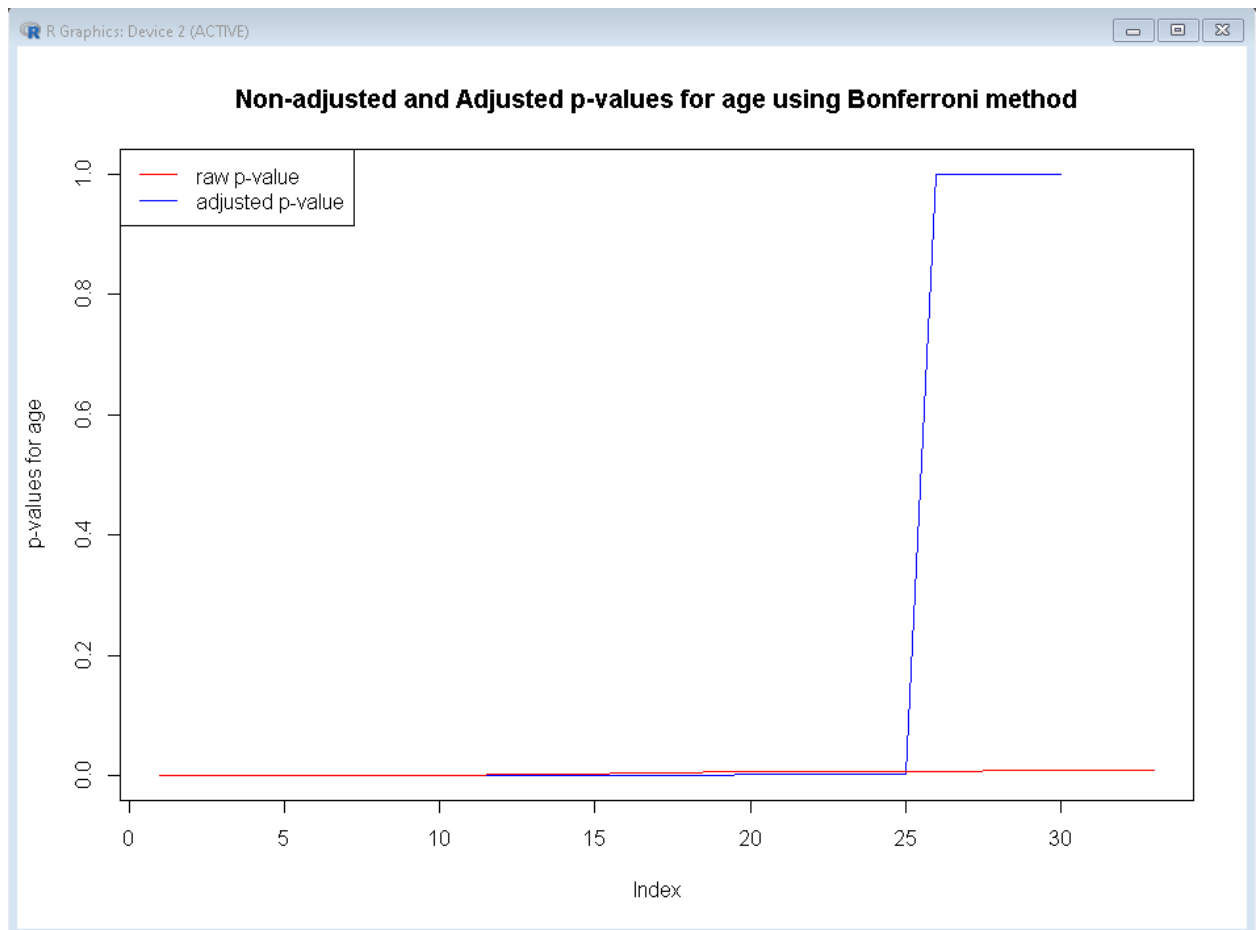
```



```

plot(sortage2, type = "l", ylab = "p-values for age", ylim = c(0,1))
lines(sortagea2, col="blue")
lines(sortage2, col="red")
title("Non-adjusted and Adjusted p-values for age using Bonferroni method")
legend("topleft", c("raw p-value", "adjusted p-value"), lty = c("solid", "solid"), col = c("red", "blue"))

```



7.

```
sam <- read.table("tcga_brca_fpk_m_sam.txt", header=T, row.names=1, fill=TRUE)
fpkm <- read.table("tcga_brca_fpk_m.txt", header=T, row.names=1)
```

```
8. gata3 <- fpkm[grepl(pattern="GATA3", rownames(fpkm)), ]
as.numeric(gata3)
```

```
9. binary <- +(gata3 >= quantile(gata3, 0.75))
```

```
[1] 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0
[37] 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 1 1 1 0 1 0 1 0 0 1 0 0 0 1 0
[73] 0 0 0 0 1 0 0 0 0 1 1 1 1 0 0 1 0 1 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[109] 0 0 0 0 0 0 0 0 1 0 1
```

```
10. group <- cbind(binary, sam)
```

11.

```
g1 <- group %>%
  rename(status = vital_status,
    time = months_to_event) %>%
  mutate(status = case_when(status=="LIVING" ~ 0,
    status=="DECEASED" ~ 1))
f <- survdiff(Surv(time, status) ~ binary, data = g1)
```

Chisq= 3.2 on 1 degrees of freedom, p= 0.07

```
> f <- survdiff(Surv(time, status) ~ binary, data = g1)
> f
Call:
survdiff(formula = Surv(time, status) ~ binary, data = g1)

n=116, 3 observations deleted due to missingness.

      N Observed Expected (O-E)^2/E (O-E)^2/V
binary=0 86      27    22.39    0.951    3.22
binary=1 30       5     9.61    2.214    3.22

Chisq= 3.2 on 1 degrees of freedom, p= 0.07
.
```

12.

```
fit <- coxph(Surv(time,status)~binary ,data =g1)
> fit <- coxph(Surv(time,status)~binary ,data =g1)
> fit
Call:
coxph(formula = Surv(time, status) ~ binary, data = g1)

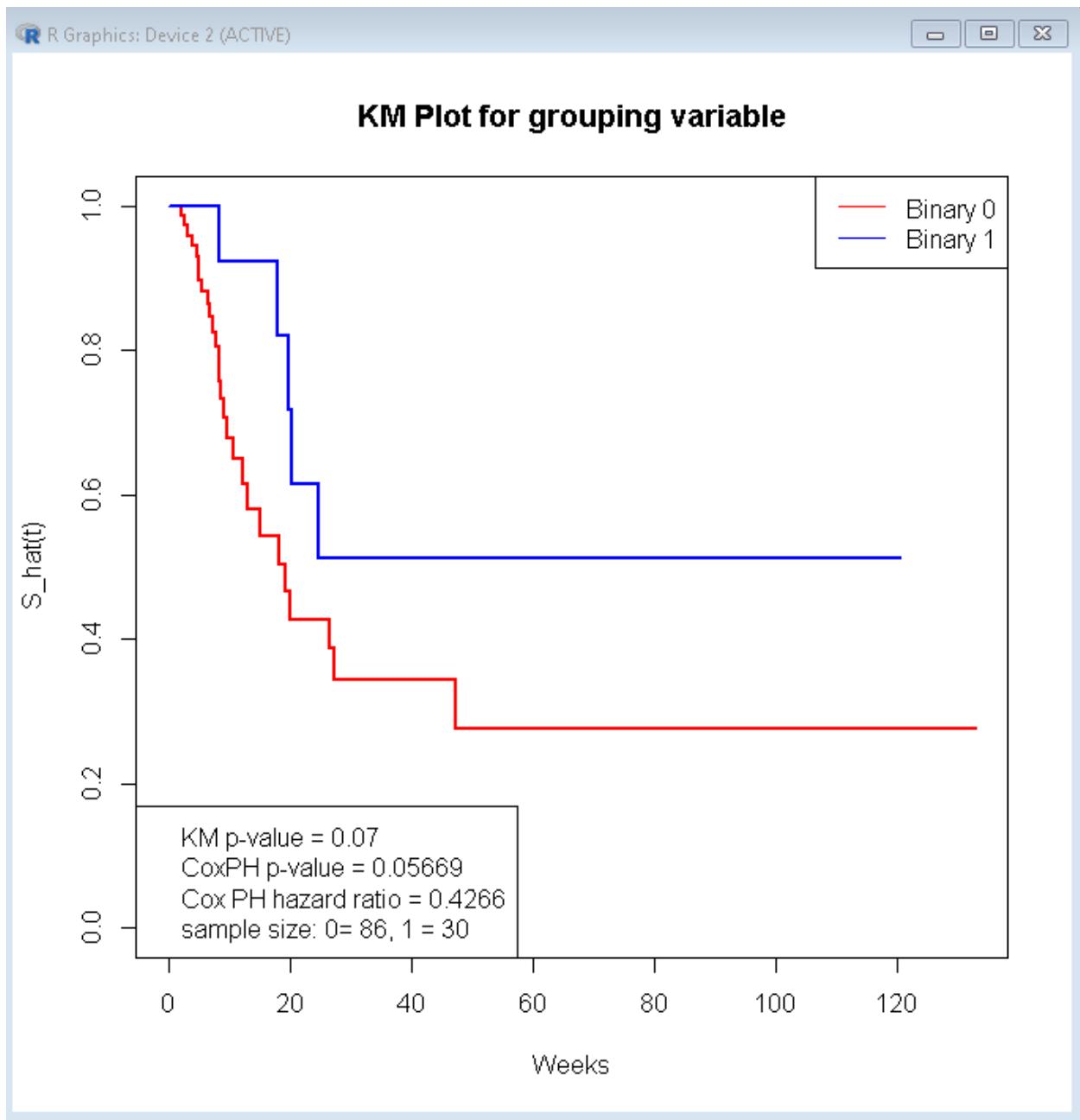
      coef exp(coef) se(coef)      z      p
binary -0.8520    0.4266   0.4893 -1.741 0.0816

Likelihood ratio test=3.63 on 1 df, p=0.05669
n= 116, number of events= 32
. (3 observations deleted due to missingness)
```

13.

```
KM <- survfit(Surv(time, status) ~ binary, data = g1)
```

```
plot(KM,lwd=2,col=c("red","blue"),xlab='Weeks',ylab='S_hat(t)',main='KM Plot for grouping variable')
legend("topright", c("Binary 0", "Binary 1"), lty = c("solid", "solid"), col = c("red", "blue"))
legend("bottomleft", c("KM p-value = 0.07", "CoxPH p-value = 0.05669", "Cox PH hazard ratio = 0.4266",
"sample size: 0= 86, 1 = 30")
```



14.

The graph produced in question 13 does not seem to depict similar results like in the study. The cox model depicted on *gat3* had a hazard ratio of .12 and a p value of 0.05 in the study which depicts a close value to what we produced in our results. Overall, the results do not really agree with what was produced in the Mehra et al study.