

Answers to Exercises

Introduction to R Graphics

1. Create a folder in which to store your R work (H drive or D drive)
2. Open R
3. Clear workspace
4. Change Dir to the folder in (1)
5. Read in the data file "thedata.txt" into a data frame called "dat". The data file is located at <http://csde.washington.edu/courses/statcore/RGraphicsJan2008/>

EXERCISE A

1. Modify the following plot by adding/using the following tags

```
cex.axis
cex.lab
cex.main
col.axis
col.lab
col.main
```

```
plot(dat$age,dat$income,xlab="Age",ylab="Income",
      main="Relationship between Age and Income",
      pch=20,col="red")
```

An answer

```
pdf(file="ExerciseA1.pdf",height=8,width=9)
plot(dat$age,dat$income,xlab="Age",ylab="Income",
      main="Relationship between Age and Income",
      pch=20,col="red",cex.axis=0.5, cex.lab=1.2, cex.main=2,
      col.axis="blue",col.lab="green",col.main="purple")
dev.off()
```

2. Make all of the changes listed below to the following code
 - (a) change the file name
 - (b) change the point character (symbol) and its color
 - (c) make the regression line purple and wider
 - (d) add a horizontal line at the median income using the `lines()` function
 - (e) change the CC region points from green to blue
 - (f) change the location of the legend and make it reflect the change in (a & e)

```
pdf(file="scatter.pdf",height=8,width=10)

plot(dat$age,dat$income,xlab="Age",ylab="Income",
     main="Relationship between Age and Income",
     pch=20,col="red")
abline(fit.age)
segments(x0=median(dat$age),y0=-5500,x1=median(dat$age),y1=15500,
         lty="dotted",lwd=2,col="blue")
points(dat$age[dat$region=="CC"],dat$income[dat$region=="CC"],
       col="green",pch=20)
legend(25,500,col=c("green","red"),pch=20,
      legend=c("CC region","All other regions"))

dev.off()
```

An answer

```
pdf(file="ExerciseA2.pdf",height=8,width=10)

plot(dat$age,dat$income,xlab="Age",ylab="Income",
     main="Relationship between Age and Income",
     pch=8,col="green")
abline(fit.age,col="purple",lwd=2)
segments(x0=median(dat$age),y0=-5500,x1=median(dat$age),y1=15500,
         lty="dotted",lwd=2,col="blue")
lines(x=c(15, 65),y=c(median(dat$income),median(dat$income)),
      lty="dotdash",lwd=2,col="orange")
points(dat$age[dat$region=="CC"],dat$income[dat$region=="CC"],
       col="blue",pch=8)
legend(45,0,col=c("blue","green"),pch=8,
      legend=c("CC region","All other regions"))

dev.off()
```

EXERCISE B

1. Run the following code

```
region <- unique(dat$region)
mean.age <- NULL
for (i in 1:length(region))
{
  mean.age <- c(mean.age, mean(dat$age[dat$region==region[i]]))
}
mean.income <- NULL
for (i in 1:length(region))
{
  mean.income <- c(mean.income,
mean(dat$income[dat$region==region[i]]))
}
```

Then make all of the changes listed below to the following code

- (a) save it as a pdf
- (b) remove the title
- (c) put the title inside the box using the text function, make it big and italic
- (d) use the `log` tag in `plot` to log the x-axis

```
plot(dat$age,dat$income,xlab="Age",ylab="Income",
     main="Relationship between Age and Income",
     pch=20,col="red")

text(x=mean.age, y=mean.income, labels=region, col="blue", cex=0.8,
     font=2)
```

An answer

```
pdf(file="ExerciseB1.pdf",height=7,width=8)
plot(dat$age,dat$income,xlab="Age",ylab="Income",
     pch=20,col="red", ylim=c(-2000,16000),log="x" )
text(x=mean.age, y=mean.income, labels=region, col="blue", cex=0.8,
     font=2)
text(20,15000,labels=" Relationship between Age and Income",
     cex=2,font=3,pos=4)
dev.off()
```

2. Modify the following code so that it has an x-axis with tick marks and labels at 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, and 65.

```
plot(dat$age,dat$income,xlab="Age", xlim=c(20,60),pch=20,
     ylim=c(-2000,14000), ylab="Income in Thousands",yaxt="n")

axis(side=2,at=seq(from=-2000,to=14000,by=2000),
     labels=seq(from=-2,to=14,by=2),las=1)
```

An answer

```
pdf(file="ExerciseB2.pdf",height=7,width=8)
plot(dat$age,dat$income,xlab="Age", xlim=c(15,65),pch=20,
     ylim=c(-2000,14000), ylab="Income in Thousands",yaxt="n",xaxt="n")
axis(side=2,at=seq(from=-2000,to=14000,by=2000),
     labels=seq(from=-2,to=14,by=2),las=1)
axis(side=1,at=seq(15,65,5))
dev.off()
```

EXERCISE C

1. Make a barplot of `table(dat$region, dat$gender)`, you decide whether to use `beside=TRUE`
 - (a) make each region a different color bar (solid not stripes)
 - (b) include a legend
 - (c) label the x-axis "Gender"
 - (d) change the x-axis labels to Female, Male, Nux from F, M, N
 - (e) save this as a Windows metafile using the `win.metafile` function and import it into a Word document (hint: use a .wmf extension and then Insert-Picture-From File)

An answer

```
win.metafile(file="ExerciseC1.wmf",height=7,width=9)
barplot(table(dat$region, dat$gender),beside=TRUE,xlab="Gender",
        names.arg=c("Female","Male","Nux"),
        col=c("red","green","blue","orange","purple"),
        legend.text=region)
dev.off()
```

2. Modify the following code:
 - (a) change the name of the file
 - (b) make the page height 8 and the page width 8
 - (c) put the plotting symbols in a horizontal line instead of a vertical line
 - (d) put the number corresponding to each plotting symbol above it
 - (e) center the plot on the page

```
pdf(file="pch.pdf",height=8,width=2)
plot(NULL,NULL,xlim=c(0,2),ylim=c(0,25),xaxt="n",yaxt="n",type="n",
     xlab="",ylab="",main="pch")
axis(side=2,at=0:25,las=1)
points(rep(1,26),0:25,pch=0:25)
dev.off()
```

An answer

```
pdf(file="ExerciseC2.pdf",height=8,width=8)
### (bottom, left, top, right)
par(mai=c(3.5,1,3.5,1))
plot(NULL,NULL,ylim=c(0,2),xlim=c(0,25),xaxt="n",yaxt="n",type="n",
     xlab="",ylab="",main="pch")
axis(side=3,at=0:25,las=1,cex.axis=0.7)
points(y=rep(1,26),x=0:25,pch=0:25)
dev.off()
```

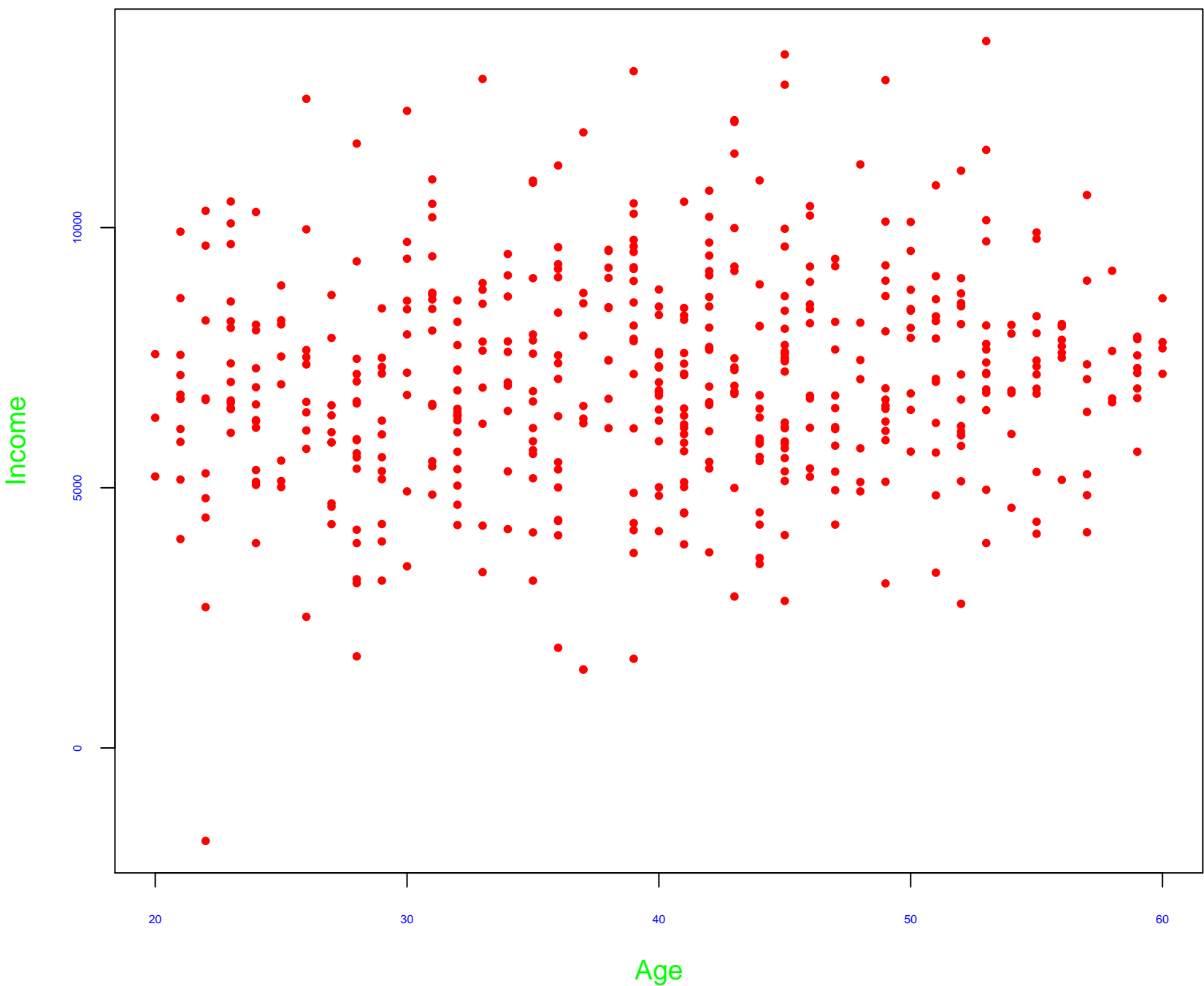
3. Sample 100 values from a normal distribution. Use the function `hist` to make a histogram of the probability densities (not frequencies). Use the function `lines` to add the normal probability density function to the histogram.

An answer

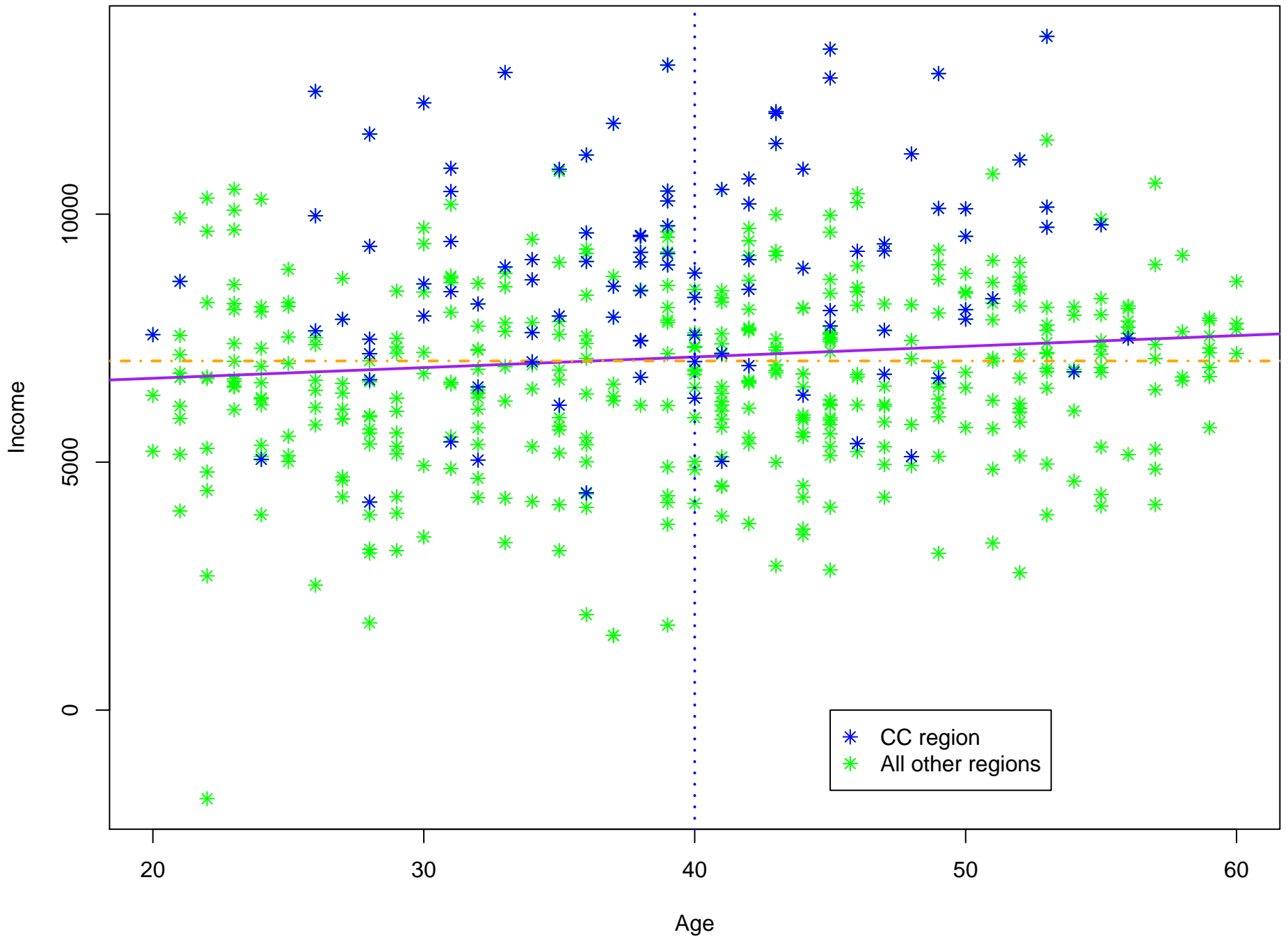
```
set.seed(222)
normal.sample <- rnorm(100)
x <- seq(-3,3,.1)
y <- dnorm(x)

pdf(file="ExerciseC3.pdf", height=8, width=10)
hist(normal.sample, freq=FALSE, breaks=seq(-3.75,3.75,.5),
      xlab="",ylab="",las=1,
      main="Random sample of 100 and its fit to the normal pdf")
lines(x,y,col="blue")
dev.off()
```

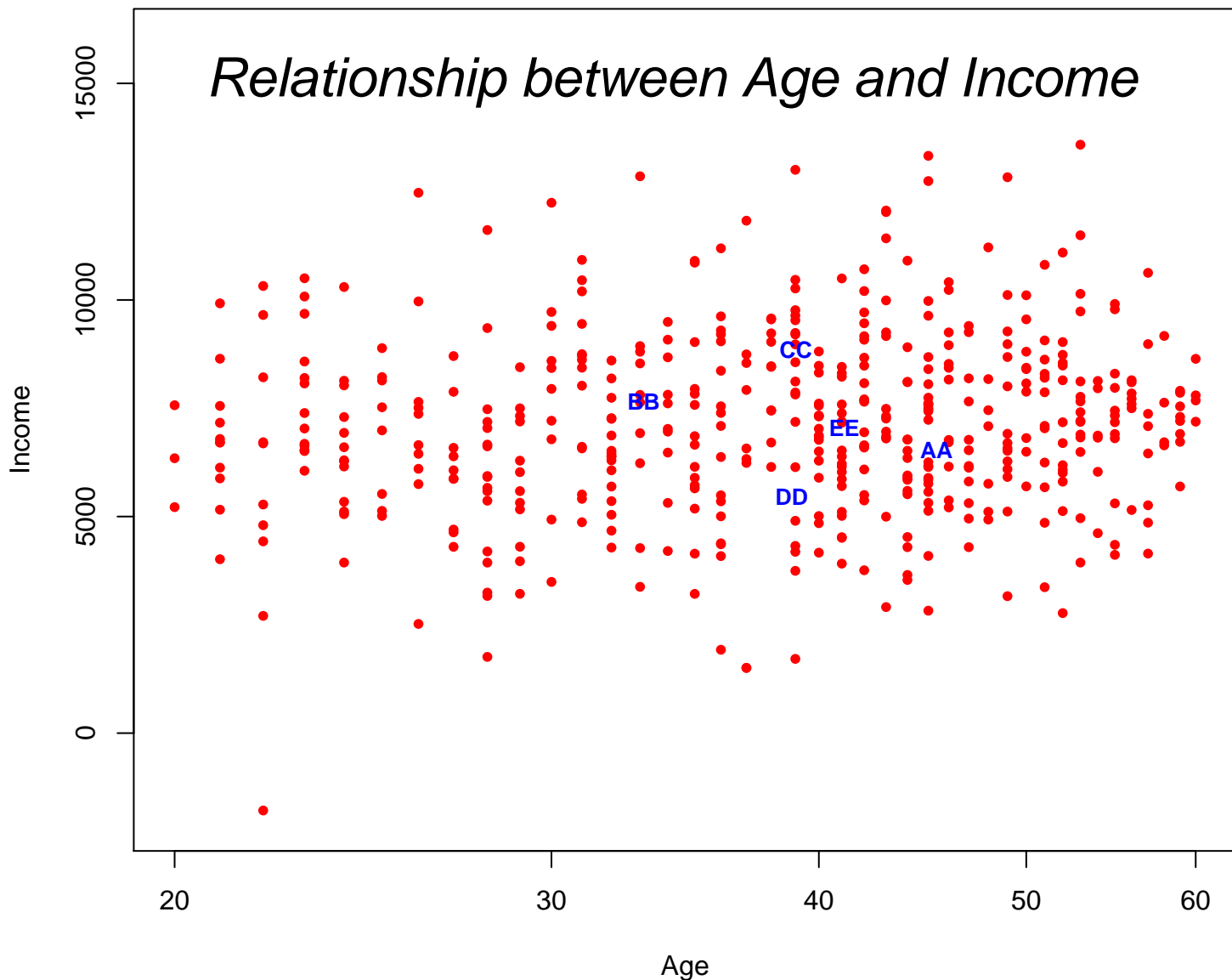
Relationship between Age and Income



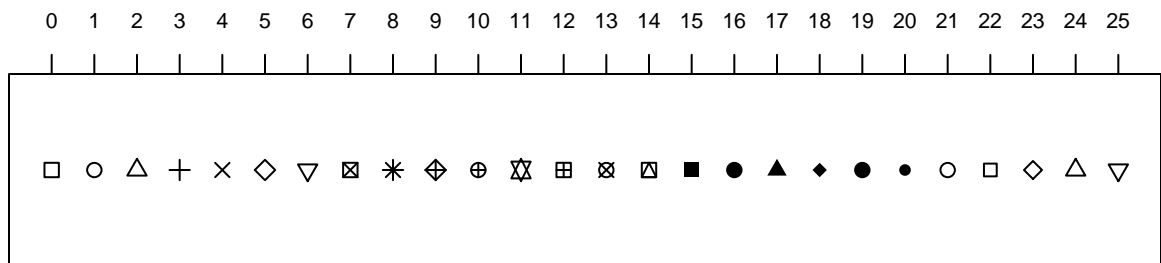
Relationship between Age and Income



Relationship between Age and Income



pch



Random sample of 100 and its fit to the normal pdf

