Math 3070 § 1. **Hockey Example: Test Two Variances** Name: Example
Treibergs **with Unequal Degrees of Freedom** June 2, 2011

Data File Used in this Analysis:

```
# Math 3070 - 1      Hockey Stick Example       June 4, 2011
# Treibergs
#
# From Navidi, "Statistics for Engineers and Scientists, 2nd. ed.,"
# McGraw Hill, 2008
#
# The breaking strength of graphite-kevlar hockey sticks is recorded.
#
# Variables
#         Strength     (newtons)
#         Composite    A or B
Strength Compsite
487.3 A
444.5 A
467.7 A
456.3 A
449.7 A
459.2 A
478.9 A
461.5 A
477.2 A
488.5 B
501.2 B
475.3 B
467.2 B
462.5 B
499.7 B
470.0 B
469.5 B
481.5 B
485.2 B
509.3 B
479.3 B
478.3 B
491.5 B
488.5 B
```

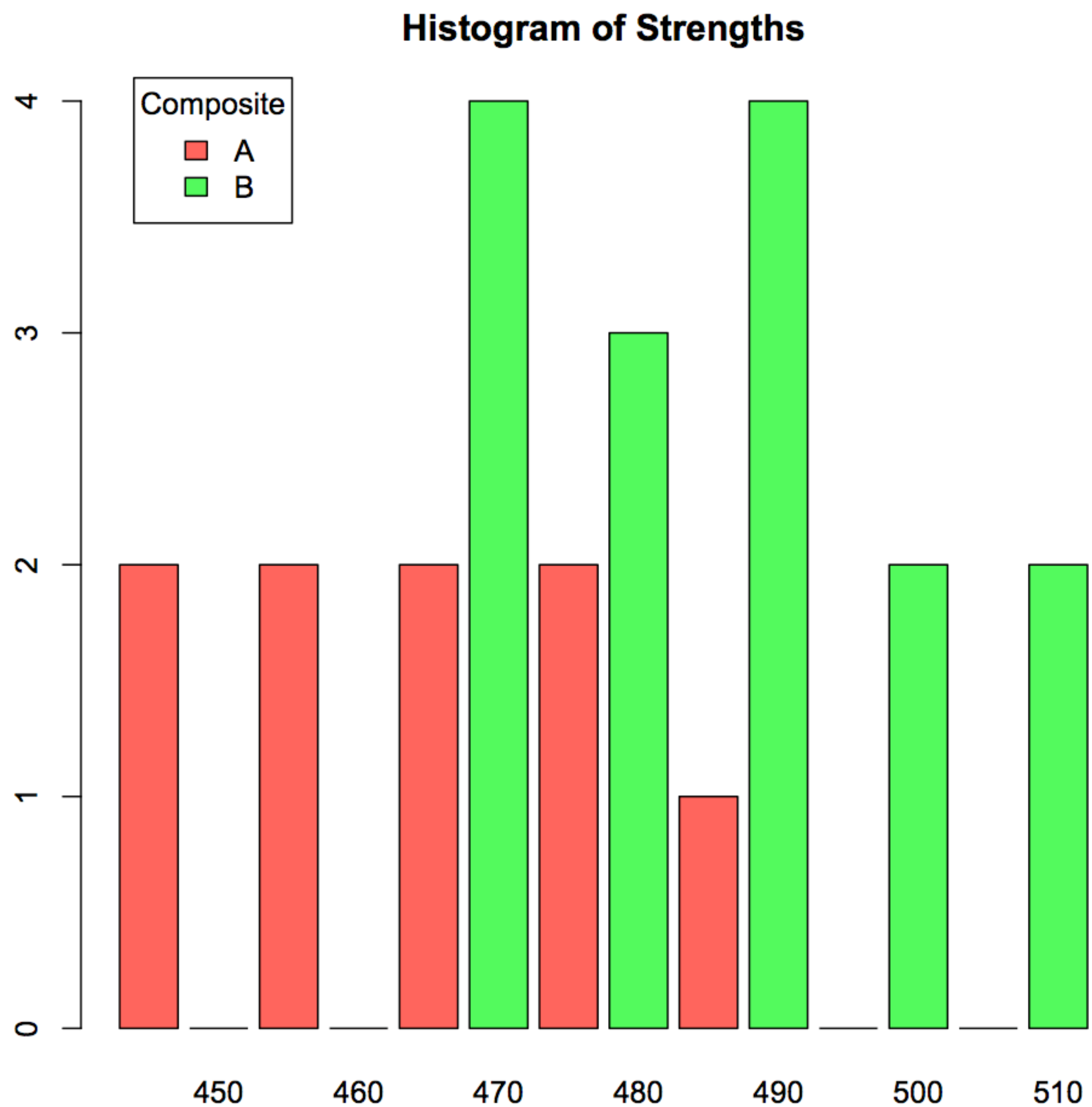
R Session:

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
 Natural language support but running in an English locale
```

```
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]
[Workspace restored from /Users/andrejstreibergs/.RData]

> tt <- read.table("M3074HockeyData.txt",header=TRUE)
> attach(tt)
> tt
   Strength Compsite
1     487.3        A
2     444.5        A
3     467.7        A
4     456.3        A
5     449.7        A
6     459.2        A
7     478.9        A
8     461.5        A
9     477.2        A
10    488.5        B
11    501.2        B
12    475.3        B
13    467.2        B
14    462.5        B
15    499.7        B
16    470.0        B
17    469.5        B
18    481.5        B
19    485.2        B
20    509.3        B
21    479.3        B
22    478.3        B
23    491.5        B
24    488.5        B
> ################# SUMMARIZE THE DATA ######################################
> tapply(Strength,Compsite,summary)
$A
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  444.5   456.3   461.5   464.7   477.2   487.3

$B
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  462.5   472.6   481.5   483.2   490.0   509.3

> tapply(Strength,Compsite,var)
       A        B
202.7175 182.6752
```

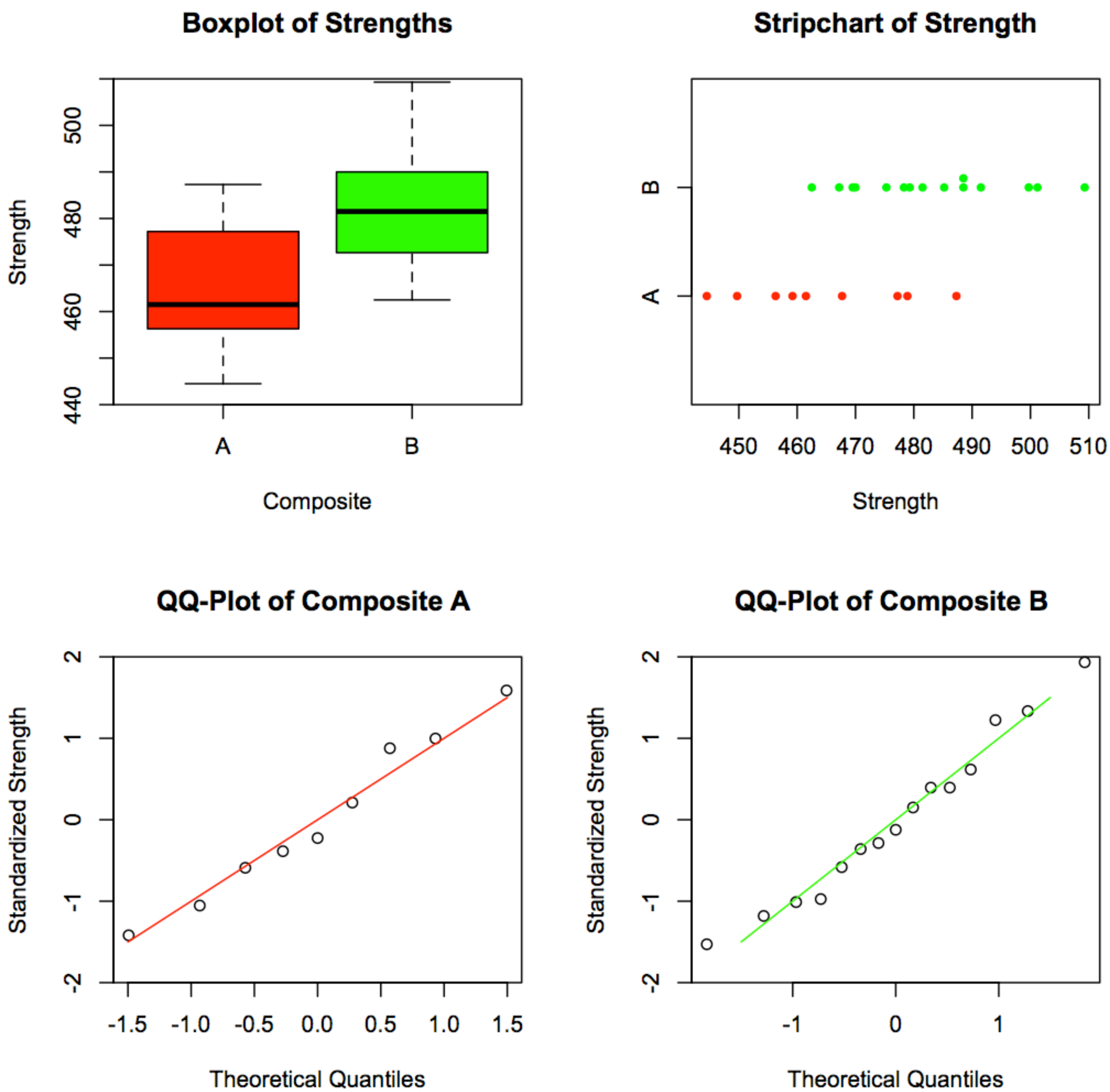
```
> # Pick off the two samples.
> y1 <- Strength[Compsite=="A"]
> y1
[1] 487.3 444.5 467.7 456.3 449.7 459.2 478.9 461.5 477.2
> y2 <- Strength[Compsite=="B"]
> y2
 [1] 488.5 501.2 475.3 467.2 462.5 499.7 470.0 469.5 481.5 485.2 509.3 479.3
[13] 478.3 491.5 488.5


> ###################### SIDE-BY-SIDE HISTOGRAM "BY HAND" ##################
> #  We shall use hist() to compute the bin counts. Then we'll use boxplot()
> # to draw the bars. Take a look at the data together.
> hist(Strength)
> # Take the same breakpoints.
> br <- seq(440,510,10)
> br
[1] 440 450 460 470 480 490 500 510

> # We can pick off number in each bin.
> hh <- hist(y1,breaks=br,ylim=c(0,6))

> # The class "histogram" is list with vectors used in plotting.
> hh
$breaks
[1] 440 450 460 470 480 490 500 510

$counts
[1] 2 2 2 2 1 0 0

$intensities
[1] 0.02222222 0.02222222 0.02222222 0.02222222 0.01111111 0.00000000 0.00000000

$density
[1] 0.02222222 0.02222222 0.02222222 0.02222222 0.01111111 0.00000000 0.00000000

$mids
[1] 445 455 465 475 485 495 505

$xname
[1] "y1"

$equidist
[1] TRUE

attr(,"class")
[1] "histogram"
```

```
> # We care about "counts"
>
> h1 <- hist(y1,breaks=br)
> h2 <- hist(y2,breaks=br)
> c1 <- h1$counts
> c2 <- h2$counts
> matr <- t(matrix(c(c1,c2),ncol=2))
> matr
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    2    2    2    2    1    0    0
[2,]    0    0    4    3    4    2    2

> nam <- rbind(rep("",7),br[2:8])
> colr <- rainbow(3,alpha=.7)[1:2]
> barplot(matr,beside=T,col=colr,space=c(0.2,0.2),names=nam)
> title("Histogram of Strengths")
> legend(.445,4.1,fill=colr,legend=c("A","B"),title="Composite")
> # M3074Hockey1.pdf
```



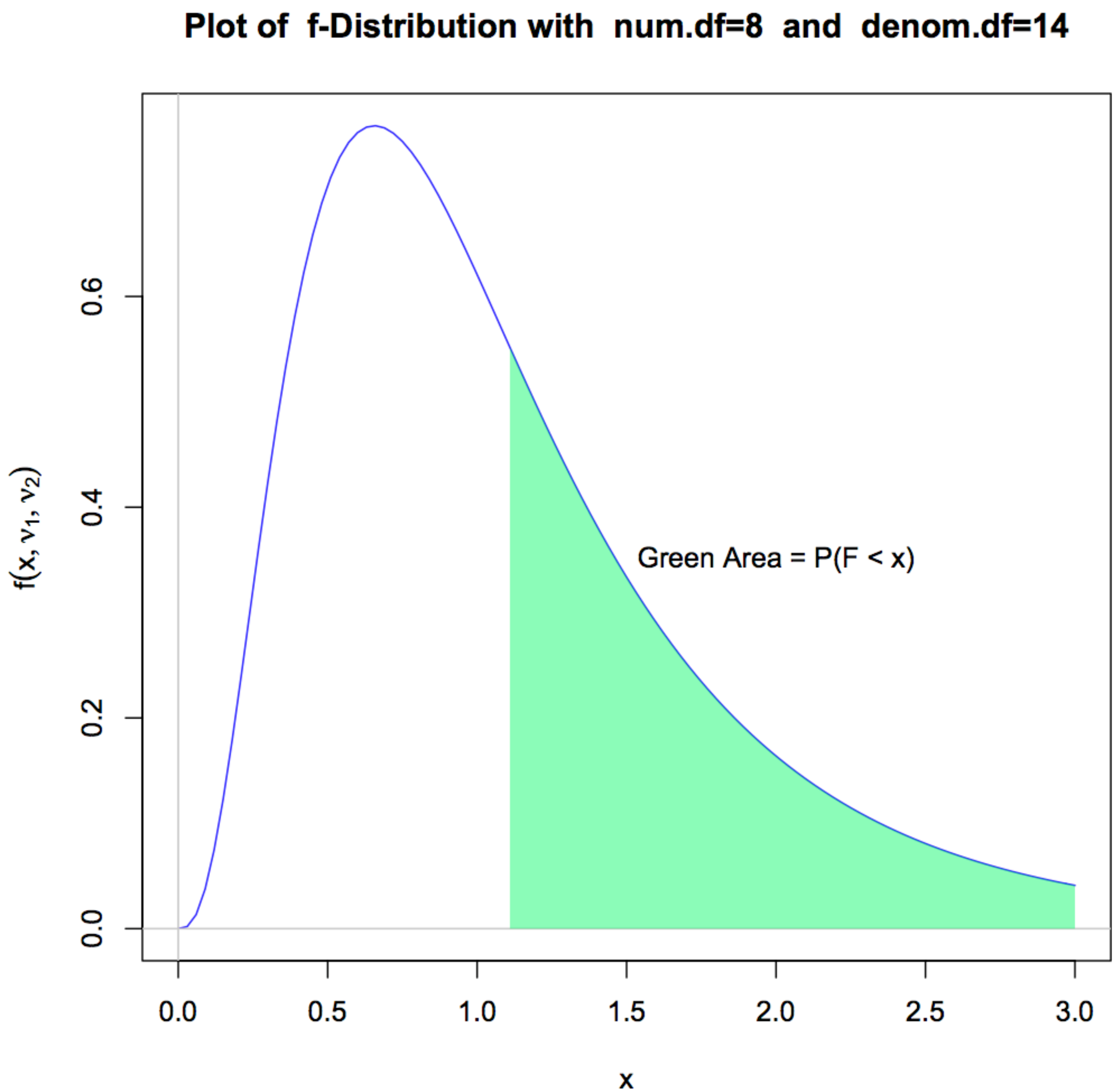
Histogram of Strengths
Composite
A
B
0
1
2
3
4
450
460
470
480
490
500
510

```
> ################## DIAGNOSTIC PLOTS ON STRENGTHS ############################
> layout(matrix(c(1,3,2,4),ncol=2)
>
> boxplot(Strength~Compsite,ylim=c(440,510),ylab="Strength",
+ xlab="Composite",main="Boxplot of Strengths",col=rainbow(4))
>
> stripchart(Strength~Compsite,col=rainbow(3),method="stack",
+ ylim=c(0,3), main="Stripchart of Strength",pch=16)
>
> qqnorm((y1-mean(y1))/sd(y1),ylab="Standardized Strength",
+ main="QQ-Plot of Composite A ",ylim=c(-2,2))
> curve(x-0,-1.5,1.5,col=rainbow(4)[1],add=T)
>
> qqnorm((y2-mean(y2))/sd(y2),ylab="Standardized Strength",
+ main="QQ-Plot of Composite B ",ylim=c(-2,2))
> curve(x-0,-1.5,1.5,col=rainbow(4)[2],add=T)
> # M3074Hockey2.pdf
>
> # The lengths of the boxes are about the same, indicating variances
> # should be nearly the same.
> # The normal QQ-Plots are linear indicating that both populations
> # are plausibly normal.
```


Boxplot of Strengths
Strength
440
460
480
500
A
B
Composite
Stripchart of Strength
B
A
450
460
470
480
490
500
510
Strength
QQ-Plot of Composite A
Standardized Strength
-2
-1
0
1
2
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Theoretical Quantiles
QQ-Plot of Composite B
Standardized Strength
-2
-1
0
1
2
-1
0
1
Theoretical Quantiles

```
>
> ##################### TEST FOR EQUALITY OF VARIANCES #####################
>
> var.test(y1,y2)

F test to compare two variances

data:  y1 and y2
F = 1.1097, num df = 8, denom df = 14, p-value = 0.8254
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3377832 4.5827526
sample estimates:
ratio of variances
          1.109715

> # High p-value: we cannot reject null hypothesis that ratio of variances = 1

> #################### TEST FOR EQUALITY OF VARIANCE "BY HAND" ##############
> n1 <- length(y1); n1
[1] 9
> n2 <- length(y2); n2
[1] 15
> F <- var(y1)/var(y2); F
[1] 1.109715

> alpha <- .05

> fhighcrit <- qf(alpha/2,n1-1,n2-1,lower.tail=F); fhighcrit
[1] 0.2421504

> fhighcrit <- qf(alpha/2,n1-1,n2-1,lower.tail=FALSE); fhighcrit
[1] 3.285288

> # CI for sigma1^2 / sigma2^2
> c(F/fhighcrit,F/flowcrit)
[1] 0.3377832 4.5827526

> # With unequal df's, critical values are not reciprocal.
> 1/flowcrit
[1] 4.129665


> # For two-tailed test the p-value is smaller area multiplied by 2
> pv <- 2*pf(F,n1-1,n2-1,lower.tail=FALSE); pv
[1] 0.8253555
```

```
> #################### PLOT  f-DISTRIBUTION WITH AREA F<X ####################
> layout=(1)
> F <- 1.109715
> xmax <- 3
> curve(df(x,n1-1,n2-1),0,xmax,ylab=expression(f(x,nu[1],nu[2])),
+ main="Plot of  f-Distribution with  num.df=8  and  denom.df=14",col=4)
> abline(h=0,col="gray"); abline(v=0,col="gray")
> x <- seq(from=F, to=xmax, by=(xmax-F)/100)
> y <- df(x,n1-1,n2-1)
> polygon(c(F,x,x[101],F),c(0,y,0,0),col=rainbow(5,alpha=.5)[3],border=NA)
text(2,.35,"Green Area = P(F < x)")
> # M3074Hockey3.pdf
```



Plot of f-Distribution with num.df=8 and denom.df=14
Green Area = P(F < x)
f(x, ν1, ν2)
x
0.0
0.2
0.4
0.6
0.0
0.5
1.0
1.5
2.0
2.5
3.0