Homework for week 7 - Data Management

Joe Brew

1. Please, run the SAS job called Ex1.ReadinDta.sas I sent you a week ago. I did review the contents of this SAS code file during our last session, but I did not execute/run it due to lack of time. Make sure you run it and send me your SAS output file in HTML or PDF format.
If you have any questions, make sure you make note of them in your email.

PDF ATTACHED TO THIS DOCUMENT

2. Please copy and paste the following lines of data then write a SAS code using "INFILE", "INPUT" and "DATALINES" statement to read it into SAS:

```
ABA,48,100,50,"Webster Groves, St. Louis",USA
Matt,52,20,150,"Saint Germain, Tunis",Tunisia;
Michael,43,200,225,"Maryland Heights, Kansas City",    USA
Alex,62,1000,2000,"Coon Rapids, Ohio", 'US of A'
Brad, 33,100,100,"O'Fallon, Illinois", USA;
Schaltenbrand,28, , ," ", Germany.
```

This data includes a name, Age, donations for 2014 & 2015, address and country name. Use "PROC PRINT" to show that SAS read this data correctly.

I DID THIS IN R (SEE ATTACHED PDF) SINCE I PREFER NON-PROPRIETARY SOFTWARE, BUT HERE'S MY SAS CODE:

```
proc import datafile='/home/joebrew/phc7065/week_7_data.csv'
    out=mydata
    dbms=csv
    replace;
    getnames=yes;
run;

proc print;
run;
```
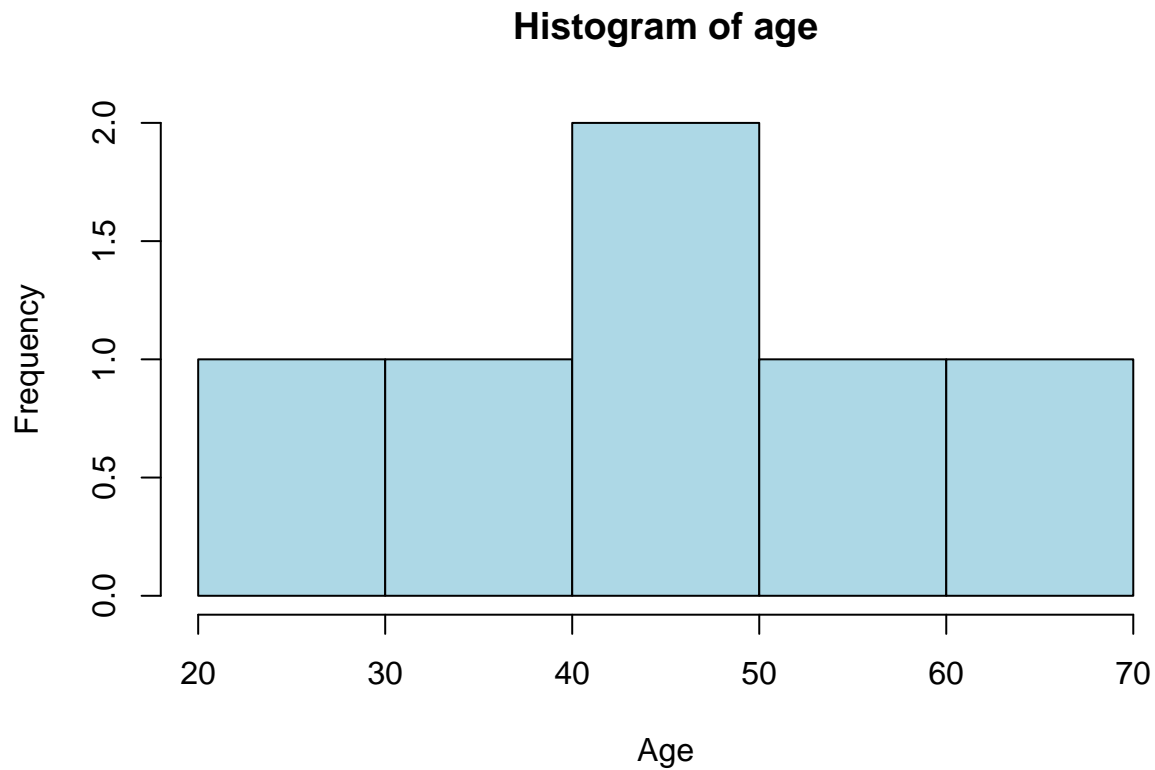
week_7.R

joebrew

Thu Feb 19 16:36:25 2015

```r
setwd('/home/joebrew/Documents/uf/phc7065')
dat <- read.csv('week_7_data.csv')

hist(dat$age, main = "Histogram of age",
     xlab = "Age", col = "lightblue")
```
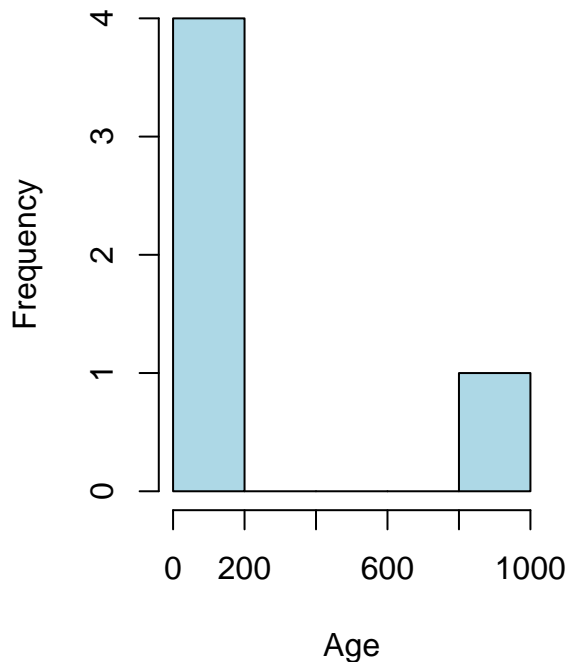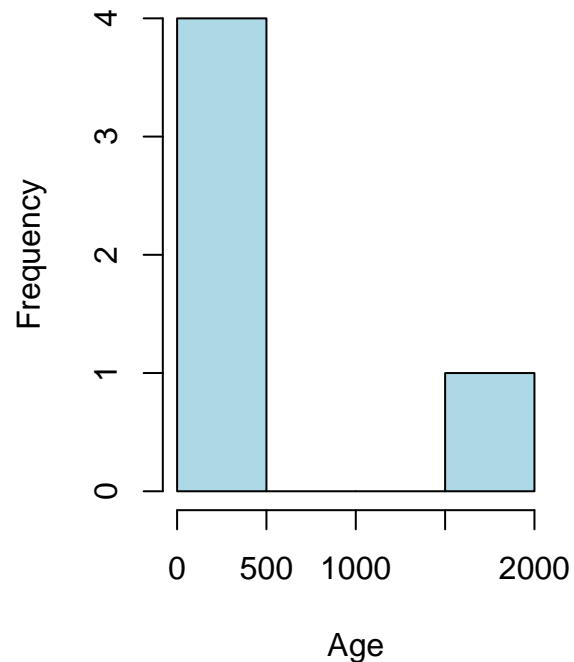
Histogram of age


```r
par(mfrow = c(1,2))
hist(dat$donations2014, main = "Histogram of 2014 donations",
     xlab = "Age", col = "lightblue")

hist(dat$donations2015, main = "Histogram of 2015 donations",
     xlab = "Age", col = "lightblue")
```

Histogram of 2014 donations

Histogram of 2015 donations


```
par(mfrow = c(1,1))

x <- data.frame(donations2014 = dat$donations2014,
                donations2015 = dat$donations2015)
scatterBarNorm <- function(x, dcol="blue", lhist=20, num.dnorm=5*lhist, ...){
  x <- x[which(!is.na(x[,1]) & !is.na(x[,2])),]
  ## check input
  stopifnot(ncol(x)==2)
  ## set up layout and graphical parameters
  layMat <- matrix(c(2,0,1,3), ncol=2, byrow=TRUE)
  layout(layMat, widths=c(5/7, 2/7), heights=c(2/7, 5/7))
  ospc <- 0.5 # outer space
  pext <- 4 # par extension down and to the left
  bspc <- 1 # space between scatter plot and bar plots
  par. <- par(mar=c(pext, pext, bspc, bspc),
              oma=rep(ospc, 4)) # plot parameters
  ## scatter plot
  plot(x, xlim=range(x[,1]), ylim=range(x[,2]), ...)
  ## 3) determine barplot and height parameter
  ## histogram (for barplot-ting the density)
  xhist <- hist(x[,1], plot=FALSE, breaks = 20)
#              breaks=seq(from=min(x[,1]), to=max(x[,1]),
#                                      length.out=lhist))
  yhist <- hist(x[,2], plot=FALSE, breaks = 20)
#              breaks=seq(from=min(x[,2]), to=max(x[,2]),
#                                      length.out=lhist)) # note: this uses probability=TRUE
  ## determine the plot range and all the things needed for the barplots and lines
  xx <- seq(min(x[,1]), max(x[,1]), length.out=num.dnorm) # evaluation points for the overlaid density
  xy <- dnorm(xx, mean=mean(x[,1]), sd=sd(x[,1])) # density points
```

2

```
  yx <- seq(min(x[,2]), max(x[,2]), length.out=num.dnorm)
  yy <- dnorm(yx, mean=mean(x[,2]), sd=sd(x[,2]))
  ## barplot and line for x (top)
  par(mar=c(0, pext, 0, 0))
  barplot(xhist$density, axes=FALSE, ylim=c(0, max(xhist$density, xy)),
          space=0, col = adjustcolor("red", alpha.f = 0.6)) # barplot
  lines(seq(from=0, to=lhist-1, length.out=num.dnorm), xy, col="red") # line
  ## barplot and line for y (right)
  par(mar=c(pext, 0, 0, 0))
  barplot(yhist$density, axes=FALSE, xlim=c(0, max(yhist$density, yy)),
          space=0,col = adjustcolor("blue", alpha.f = 0.6), horiz=TRUE) # barplot
  lines(yy, seq(from=0, to=lhist-1, length.out=num.dnorm), col="blue") # line
  ## restore parameters
  par(par.)
}

require(mvtnorm)
```
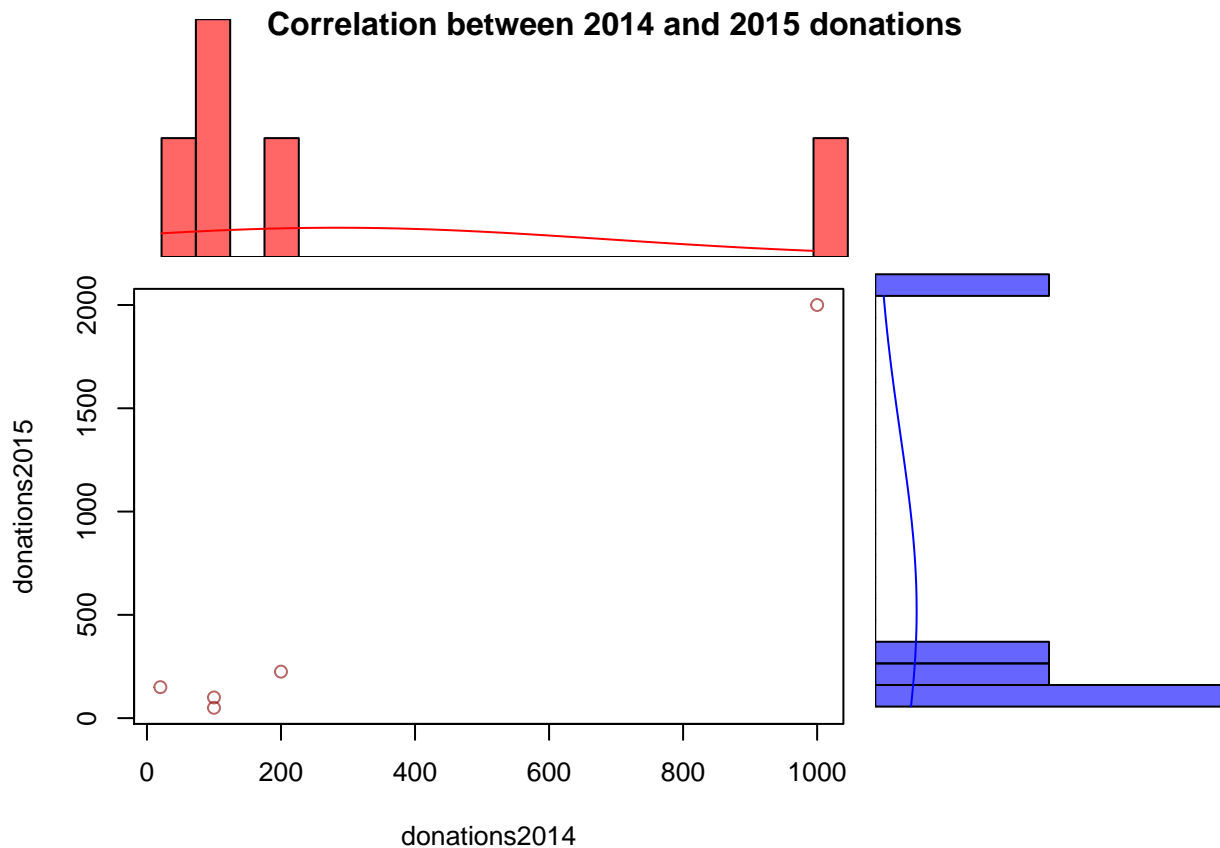
```
## Loading required package: mvtnorm
```

```
scatterBarNorm(x, col = adjustcolor("darkred", alpha.f = 0.6))
title(main = "Correlation between 2014 and 2015 donations", outer = TRUE, line = -1)

library(ggmap)
```

```
## Loading required package: ggplot2
```

```
dat$country <- gsub("[.]| of ", "", dat$country)
temp <- geocode(as.character(paste0(dat$address, ", ",dat$country)))
```

```
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=Webster+Groves,+St.+
## Google Maps API Terms of Service : http://developers.google.com/maps/terms
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=Saint+Germain,+Tunis
## Google Maps API Terms of Service : http://developers.google.com/maps/terms
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=Maryland+Heights,+Ka
## Google Maps API Terms of Service : http://developers.google.com/maps/terms
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=Coon+Rapids,+Ohio,++
## Google Maps API Terms of Service : http://developers.google.com/maps/terms
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=O'Fallon,+Illinois,+
## Google Maps API Terms of Service : http://developers.google.com/maps/terms
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=+,++Germany&sensor=f
## Google Maps API Terms of Service : http://developers.google.com/maps/terms
```

```
library(maps)
par(mar = c(0,0,0,0))
par(oma = c(0,0,0,0))
par(mfrow = c(1,1))
map("world", fill = TRUE, col = "grey", border = FALSE)
points(temp$lon, temp$lat, col = adjustcolor("darkred", alpha.f = 0.8),
       pch = 17, cex = 2)
title(main = "Location of observations in the dataset")
```


Location of observations in the dataset
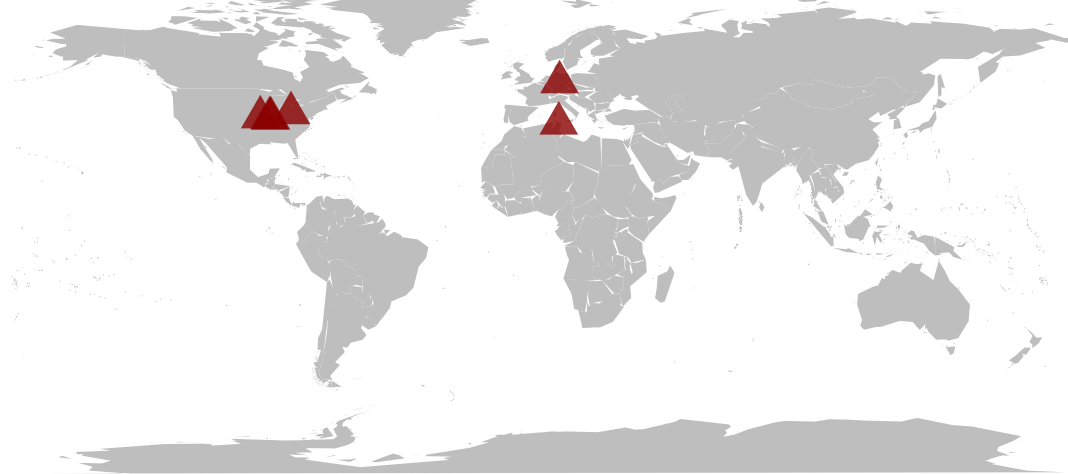
Obs	EFName	Female	Age	Age2	Salary12	Salary13	NSalary
1	Mary	Female	42.0837	42	$38,210.00	$39,547.35	$39,547.00
2	John	Male	42.3940	42	$29,755.00	$30,647.65	$30,648.00
3	Beth	Female	43.8765	44	$27,985.00	$28,684.63	$28,685.00
4	Steve	Male	45.4029	45	$51,587.00	$53,650.48	$53,650.00
5	Sue	Female	45.0075	45	$29,855.00	$30,899.93	$30,900.00
6	Molly	Female	54.3665	54	$70,998.00	$72,772.95	$72,773.00

Obs	EFName	Female	Age	Age2	Salary12	Salary13	NSalary

The MEANS Procedure

Variable	N	N Miss	Mean	Median	Minimum	Maximum	Std Dev	Std Error	Variance	Skewness	Kurtosis
Female	6	0	0.33	0.00	0.00	1.00	0.52	0.21	0.27	0.97	-1.88
Salary12	6	0	41398.33	34032.50	27985.00	70998.00	16963.77	6925.43	287769586	1.32	0.91
NSalary	6	0	42700.50	35223.50	28685.00	72773.00	17407.70	7106.66	303028033	1.27	0.70
Draise	6	0	1302.17	1191.00	700.00	2063.00	529.11	216.01	279961.77	0.49	-1.35
Bonus13	6	0	3.17	3.25	2.50	4.00	0.61	0.25	0.37	0.08	-1.55

The TTEST Procedure

Variable: NSalary

Female	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female	4	42976.3	20409.8	10204.9	28685.0	72773.0
Male	2	42149.0	16264.9	11501.0	30648.0	53650.0
Diff (1-2)		827.3	19456.5	16849.9		

Female	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		42976.3	10499.7	75452.8	20409.8	11561.9	76099.0
Male		42149.0	-103985	188283	16264.9	7256.6	519015
Diff (1-2)	Pooled	827.3	-45955.5	47610.0	19456.5	11657.1	55909.4
Diff (1-2)	Satterthwaite	827.3	-52006.0	53660.5			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	4	0.05	0.9632
Satterthwaite	Unequal	2.6475	0.05	0.9609

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	3	1	1.57	1.0000

The TTEST Procedure

Variable: NSalary
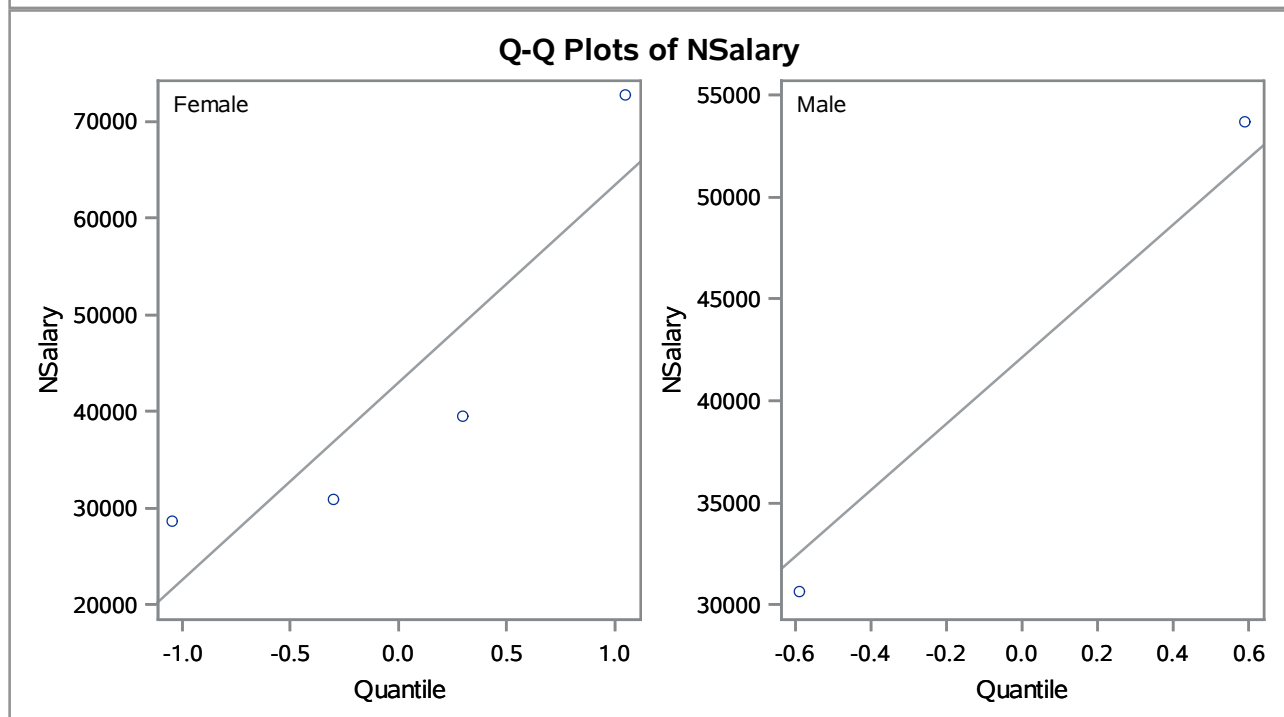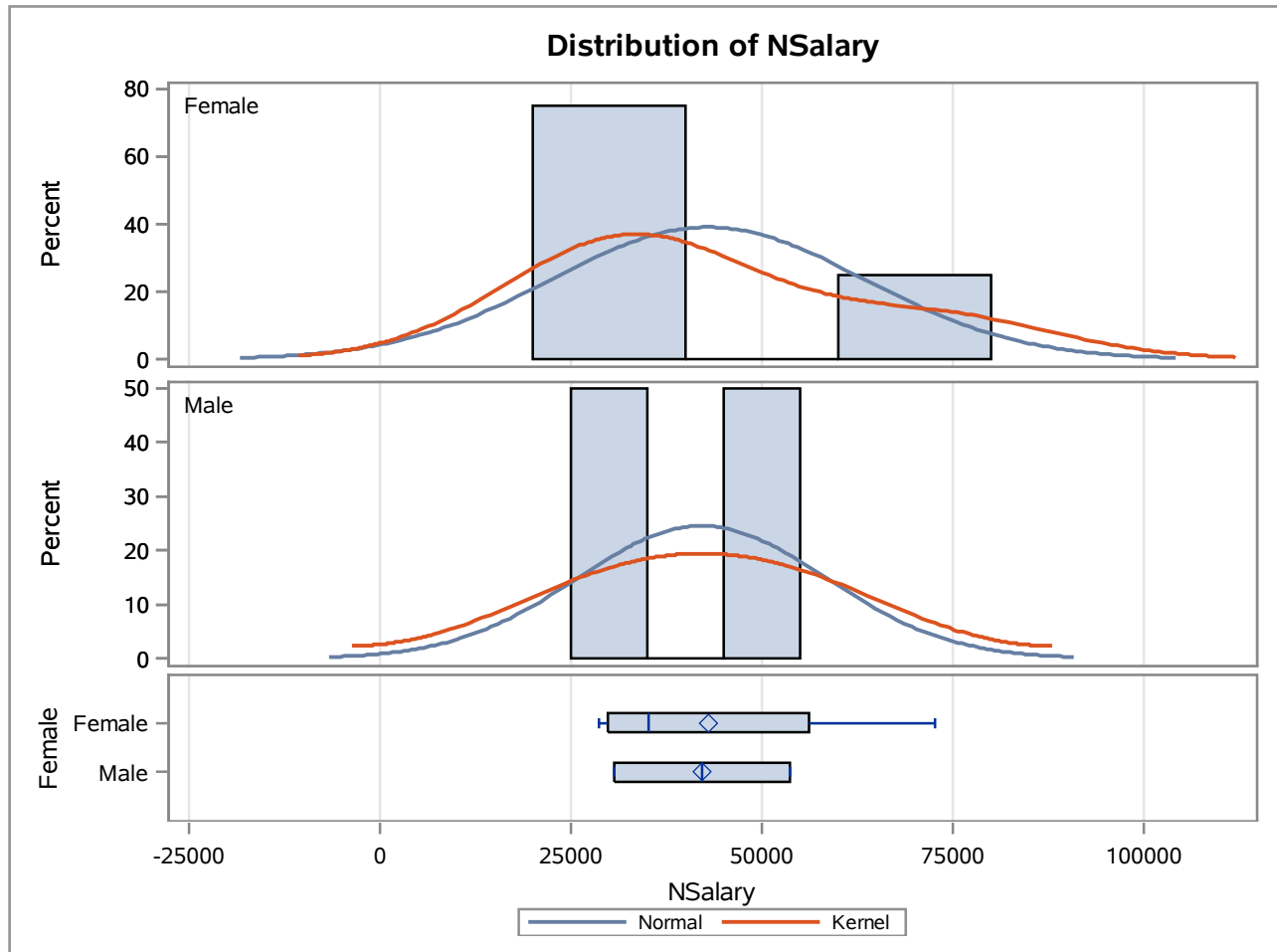
The CONTENTS Procedure

Data Set Name	WORK.PBUDGET3	Observations	0
Member Type	DATA	Variables	3
Engine	V9	Indexes	0
Created	02/19/2015 21:05:45	Observation Length	24
Last Modified	02/19/2015 21:05:45	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	5431
Obs in First Data Page	0
Number of Data Set Repairs	0
Filename	/saswork/SAS_work3A3D00006ACA_odaws03-us.oda.sas.com/SAS_work45A000006ACA_odaws03-us.oda.sas.com/pbudget3.sas7bdat
Release Created	9.0401M2
Host Created	Linux
Inode Number	30408721
Access Permission	rw-r--r--
Owner Name	joebrew
File Size (bytes)	262144

Variables in Creation Order			
#	Variable	Type	Len
1	EFName	Char	8
2	Salary12	Num	8
3	Bonus13	Num	8

The CONTENTS Procedure

Data Set Name	WORK.EMAILSURVEYDATA	**Observations**	0
Member Type	DATA	**Variables**	0
Engine	V9	**Indexes**	0
Created	02/19/2015 21:05:45	**Observation Length**	0
Last Modified	02/19/2015 21:05:45	**Deleted Observations**	0
Protection		**Compressed**	NO
Data Set Type		**Sorted**	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	74752
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	65535
Obs in First Data Page	0
Number of Data Set Repairs	0
Filename	/saswork/SAS_work3A3D00006ACA_odaws03-us.oda.sas.com/SAS_work45A000006ACA_odaws03-us.oda.sas.com/ emailsurveydata.sas7bdat
Release Created	9.0401M2
Host Created	Linux
Inode Number	30408723
Access Permission	rw-r--r--
Owner Name	joebrew
File Size (bytes)	149504

The CONTENTS Procedure

Data Set Name	WORK.TESTDATA	Observations	0
Member Type	DATA	Variables	0
Engine	V9	Indexes	0
Created	02/19/2015 21:05:45	Observation Length	0
Last Modified	02/19/2015 21:05:45	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	74752
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	65535
Obs in First Data Page	0
Number of Data Set Repairs	0
Filename	/saswork/SAS_work3A3D00006ACA_odaws03-us.oda.sas.com/SAS_work45A000006ACA_odaws03-us.oda.sas.com/testdata.sas7bdat
Release Created	9.0401M2
Host Created	Linux
Inode Number	30408725
Access Permission	rw-r--r--
Owner Name	joebrew
File Size (bytes)	149504

The CONTENTS Procedure

Data Set Name	WORK.PBUDGET2	**Observations**	6
Member Type	DATA	**Variables**	15
Engine	V9	**Indexes**	0
Created	02/19/2015 21:05:45	**Observation Length**	120
Last Modified	02/19/2015 21:05:45	**Deleted Observations**	0
Protection	READ/ALTER	**Compressed**	NO
Data Set Type		**Sorted**	YES
Encrypted	YES		
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1090
Obs in First Data Page	6
Number of Data Set Repairs	0
Filename	/saswork/SAS_work3A3D00006ACA_odaws03-us.oda.sas.com/SAS_work45A000006ACA_odaws03-us.oda.sas.com/pbudget2.sas7bdat
Release Created	9.0401M2
Host Created	Linux
Inode Number	30408727
Access Permission	rw-r--r--
Owner Name	joebrew
File Size (bytes)	262144

Alphabetic List of Variables and Attributes			
#	**Variable**	**Type**	**Len**
13	Age	Num	8
14	Age2	Num	8
6	Birth	Num	8
3	Bonus13	Num	8
10	Draise	Num	8
1	EFName	Char	8
11	Female	Num	8

The CONTENTS Procedure

#	Variable	Type	Len
Alphabetic List of Variables and Attributes			
15	HasKids	Num	8
4	Married	Char	8
9	NSalary	Num	8
5	NumofKids	Num	8
2	Salary12	Num	8
8	Salary13	Num	8
7	hired	Num	8
12	now	Num	8

Sort Information	
Sortedby	Female
Validated	YES
Character Set	ASCII