```
> library (pastecs)
> library(ggplot2)
> library(readr)
> ACS14 <- read_csv("~/GitHub/dsc520/data/acs-14-1yr-s0201.csv")

-- Column specification --------------------------------------------------------
cols(
  Id = col_character(),
  Id2 = col_double(),
  Geography = col_character(),
  PopGroupID = col_double(),
  `POPGROUP.display-label` = col_character(),
  RacesReported = col_double(),
  HSDegree = col_double(),
  BachDegree = col_double()
)

> View(ACS14)
> ##i.
> class(ACS14$Id)
[1] "character"
> class(ACS14$Id2)
[1] "numeric"
> class(ACS14$Geography)
[1] "character"
> class(ACS14$PopGroupID)
[1] "numeric"
> class(ACS14$`POPGROUP.display-label`)
[1] "character"
> class(ACS14$RacesReported)
[1] "numeric"
> class(ACS14$HSDegree)
[1] "numeric"
> class(ACS14$BachDegree)
[1] "numeric"
> ##ii.
> str(ACS14)
spec_tbl_df [136 x 8] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ Id                    : chr [1:136] "0500000US01073" "0500000US04013" "0500000US04019"
"0500000US06001" ...
 $ Id2                   : num [1:136] 1073 4013 4019 6001 6013 ...
 $ Geography             : chr [1:136] "Jefferson County, Alabama" "Maricopa County, Arizona"
"Pima County, Arizona" "Alameda County, California" ...
 $ PopGroupID            : num [1:136] 1 1 1 1 1 1 1 1 1 1 ...
```

```
 $ POPGROUP.display-label: chr [1:136] "Total population" "Total population" "Total population" "Total population" ...
 $ RacesReported        : num [1:136] 660793 4087191 1004516 1610921 1111339 ...
 $ HSDegree             : num [1:136] 89.1 86.8 88 86.9 88.8 73.6 74.5 77.5 84.6 80.6 ...
 $ BachDegree           : num [1:136] 30.5 30.2 30.8 42.8 39.7 19.7 15.4 30.3 38 20.7 ...
 - attr(*, "spec")=
  .. cols(
  ..   Id = col_character(),
  ..   Id2 = col_double(),
  ..   Geography = col_character(),
  ..   PopGroupID = col_double(),
  ..   `POPGROUP.display-label` = col_character(),
  ..   RacesReported = col_double(),
  ..   HSDegree = col_double(),
  ..   BachDegree = col_double()
  .. )
> nrow(ACS14)
[1] 136
> ncol(ACS14)
[1] 8
> ##iii.
> ACSHistogram <- ggplot(ACS14, aes(HSDegree)) + theme(legend.position = "none") + geom_histogram(aes(y = ..density..), binwidth = .5, colour = "black", fill = "white") + labs(title = "County Populations with HS Degrees", x = "% of Population with HS Degrees", y = "Density")
> ACSHistogram
> ##iv.
> ##1) Based on the histogram, the data distribution is unimodal, and the mode is 89.1%.
> getmode <- function(v) {
+   uniqv <- unique(v)
+   uniqv[which.max(tabulate(match(v, uniqv)))]
+ }
> getmode(ACS14$HSDegree)
[1] 89.1
> ##2) The histogram is not symmetrical as there some outliers and it skewed toward higher percentages of the population possessing a high school degree.
> ##3) If you do not consider the outliers, the histogram is relatively bell-shaped. If you consider the outliers, it is not.
> ##4) The histogram is not normal.
> shapiro.test(ACS14$HSDegree)

        Shapiro-Wilk normality test

data:  ACS14$HSDegree
W = 0.87736, p-value = 3.194e-09
```

```
> ### The p-value is 3.194e-09 (much smaller than .05), which indicates the distribution is not normal.
> ##5) The histogram is negatively skewed because the data skews toward the right on the graph.
> ##6)
> ACSHistogram + stat_function(fun = dnorm, args = list(mean = mean(ACS14$HSDegree, na.rm = TRUE), sd = sd(ACS14$HSDegree, na.rm = TRUE)), colour="blue", size=1)
> ##v.
> qqplotHSDegree <- qplot(sample = ACS14$HSDegree)
> qqplotHSDegree
> ##vii.
> round(stat.desc(ACS14$HSDegree, basic = FALSE, norm = TRUE), digits = 3)
      median         mean      SE.mean CI.mean.0.95          var      std.dev
      88.700       87.632        0.439        0.868       26.193        5.118
    coef.var     skewness     skew.2SE     kurtosis     kurt.2SE   normtest.W
       0.058       -1.675       -4.030        4.353        5.274        0.877
  normtest.p
       0.000
> local({r <- getOption("repos")
+ r["CRAN"] <- "https://github.com/cran"
+ options(repos=r)})
>
```