The results below are generated from an R script.

```
# Load all the libraries we need.
library(tidyverse)
library(stringr)
library(rpart)
library(partykit)
library(randomForest)
library(class)

knitr::opts_chunk$set(echo = TRUE, warning = FALSE, eval = FALSE)

setwd("/Users/hercule/Desktop/DataScience_2020/2019Spring/Intro_to_DS/final project")
# Load the data set. Assign it to the variable 'fna'.
fna <- read_csv("FNA_cancer.csv")

## Warning:  Missing column names filled in:  'X33' [33]
## Parsed with column specification:
## cols(
##  .default = col_double(),
##  diagnosis = col_character(),
##  X33 = col_character()
## )
## See spec(...)  for full column specifications.
## Warning:  569 parsing failures.
## row col  expected    actual            file
##  1  - 33 columns 32 columns 'FNA_cancer.csv'
##  2  - 33 columns 32 columns 'FNA_cancer.csv'
##  3  - 33 columns 32 columns 'FNA_cancer.csv'
##  4  - 33 columns 32 columns 'FNA_cancer.csv'
##  5  - 33 columns 32 columns 'FNA_cancer.csv'
## ...  ...  ..........  ..........  .................
## See problems(...)  for more details.

glimpse(fna)

## Observations: 569
## Variables: 33
## $ id                      <dbl> 842302, 842517, 84300903, 84348301, 84358402, 843786,...
## $ diagnosis               <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "M", "M"...
## $ radius_mean             <dbl> 17.990, 20.570, 19.690, 11.420, 20.290, 12.450, 18.25...
## $ texture_mean            <dbl> 10.38, 17.77, 21.25, 20.38, 14.34, 15.70, 19.98, 20.8...
## $ perimeter_mean          <dbl> 122.80, 132.90, 130.00, 77.58, 135.10, 82.57, 119.60,...
## $ area_mean               <dbl> 1001.0, 1326.0, 1203.0, 386.1, 1297.0, 477.1, 1040.0,...
## $ smoothness_mean         <dbl> 0.11840, 0.08474, 0.10960, 0.14250, 0.10030, 0.12780,...
## $ compactness_mean        <dbl> 0.27760, 0.07864, 0.15990, 0.28390, 0.13280, 0.17000,...
## $ concavity_mean          <dbl> 0.30010, 0.08690, 0.19740, 0.24140, 0.19800, 0.15780,...
## $ `concave points_mean`   <dbl> 0.14710, 0.07017, 0.12790, 0.10520, 0.10430, 0.08089,...
## $ symmetry_mean           <dbl> 0.2419, 0.1812, 0.2069, 0.2597, 0.1809, 0.2087, 0.179...
## $ fractal_dimension_mean  <dbl> 0.07871, 0.05667, 0.05999, 0.09744, 0.05883, 0.07613,...
## $ radius_se               <dbl> 1.0950, 0.5435, 0.7456, 0.4956, 0.7572, 0.3345, 0.446...
## $ texture_se              <dbl> 0.9053, 0.7339, 0.7869, 1.1560, 0.7813, 0.8902, 0.773...
## $ perimeter_se            <dbl> 8.589, 3.398, 4.585, 3.445, 5.438, 2.217, 3.180, 3.85...
## $ area_se                 <dbl> 153.40, 74.08, 94.03, 27.23, 94.44, 27.19, 53.91, 50....
## $ smoothness_se           <dbl> 0.006399, 0.005225, 0.006150, 0.009110, 0.011490, 0.0...
```

```
## $ compactness_se          <dbl> 0.049040, 0.013080, 0.040060, 0.074580, 0.024610, 0.0...
## $ concavity_se            <dbl> 0.05373, 0.01860, 0.03832, 0.05661, 0.05688, 0.03672,...
## $ `concave points_se`     <dbl> 0.015870, 0.013400, 0.020580, 0.018670, 0.018850, 0.0...
## $ symmetry_se             <dbl> 0.03003, 0.01389, 0.02250, 0.05963, 0.01756, 0.02165,...
## $ fractal_dimension_se    <dbl> 0.006193, 0.003532, 0.004571, 0.009208, 0.005115, 0.0...
## $ radius_worst            <dbl> 25.38, 24.99, 23.57, 14.91, 22.54, 15.47, 22.88, 17.0...
## $ texture_worst           <dbl> 17.33, 23.41, 25.53, 26.50, 16.67, 23.75, 27.66, 28.1...
## $ perimeter_worst         <dbl> 184.60, 158.80, 152.50, 98.87, 152.20, 103.40, 153.20...
## $ area_worst              <dbl> 2019.0, 1956.0, 1709.0, 567.7, 1575.0, 741.6, 1606.0,...
## $ smoothness_worst        <dbl> 0.1622, 0.1238, 0.1444, 0.2098, 0.1374, 0.1791, 0.144...
## $ compactness_worst       <dbl> 0.6656, 0.1866, 0.4245, 0.8663, 0.2050, 0.5249, 0.257...
## $ concavity_worst         <dbl> 0.71190, 0.24160, 0.45040, 0.68690, 0.40000, 0.53550,...
## $ `concave points_worst`  <dbl> 0.26540, 0.18600, 0.24300, 0.25750, 0.16250, 0.17410,...
## $ symmetry_worst          <dbl> 0.4601, 0.2750, 0.3613, 0.6638, 0.2364, 0.3985, 0.306...
## $ fractal_dimension_worst <dbl> 0.11890, 0.08902, 0.08758, 0.17300, 0.07678, 0.12440,...
## $ X33                     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```

```
# Make a copy of the original data set and delete the first and last column which are useless. Assign it
fna_new <- data.frame(fna)
fna_new <- fna_new %>% dplyr::select(-id, -X33)
glimpse(fna_new)
```

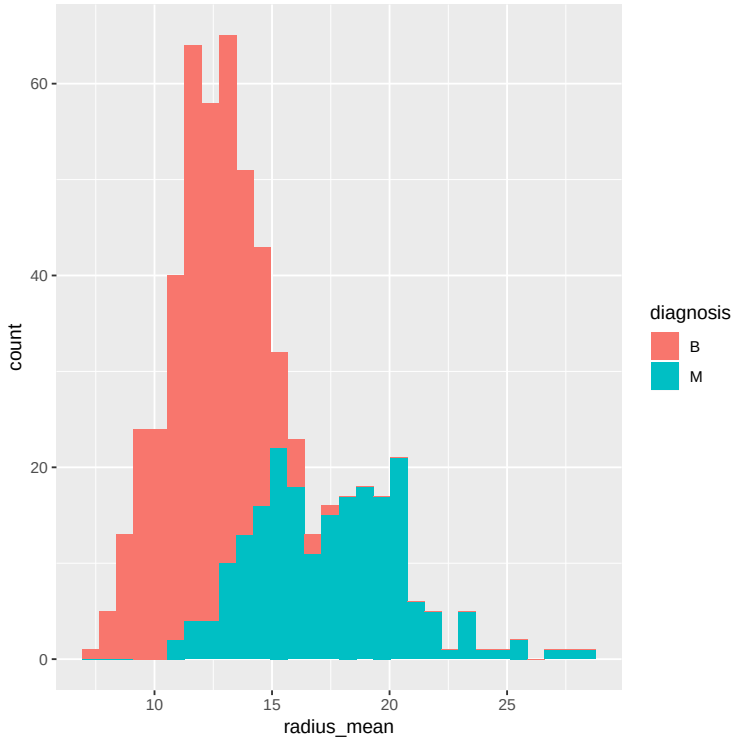
```
## Observations: 569
## Variables: 31
## $ diagnosis               <chr> "M", "M", "M", "M", "M", "M", "M", "M", "M", "M", "M"...
## $ radius_mean             <dbl> 17.990, 20.570, 19.690, 11.420, 20.290, 12.450, 18.25...
## $ texture_mean            <dbl> 10.38, 17.77, 21.25, 20.38, 14.34, 15.70, 19.98, 20.8...
## $ perimeter_mean          <dbl> 122.80, 132.90, 130.00, 77.58, 135.10, 82.57, 119.60,...
## $ area_mean               <dbl> 1001.0, 1326.0, 1203.0, 386.1, 1297.0, 477.1, 1040.0,...
## $ smoothness_mean         <dbl> 0.11840, 0.08474, 0.10960, 0.14250, 0.10030, 0.12780,...
## $ compactness_mean        <dbl> 0.27760, 0.07864, 0.15990, 0.28390, 0.13280, 0.17000,...
## $ concavity_mean          <dbl> 0.30010, 0.08690, 0.19740, 0.24140, 0.19800, 0.15780,...
## $ concave.points_mean     <dbl> 0.14710, 0.07017, 0.12790, 0.10520, 0.10430, 0.08089,...
## $ symmetry_mean           <dbl> 0.2419, 0.1812, 0.2069, 0.2597, 0.1809, 0.2087, 0.179...
## $ fractal_dimension_mean  <dbl> 0.07871, 0.05667, 0.05999, 0.09744, 0.05883, 0.07613,...
## $ radius_se               <dbl> 1.0950, 0.5435, 0.7456, 0.4956, 0.7572, 0.3345, 0.446...
## $ texture_se              <dbl> 0.9053, 0.7339, 0.7869, 1.1560, 0.7813, 0.8902, 0.773...
## $ perimeter_se            <dbl> 8.589, 3.398, 4.585, 3.445, 5.438, 2.217, 3.180, 3.85...
## $ area_se                 <dbl> 153.40, 74.08, 94.03, 27.23, 94.44, 27.19, 53.91, 50....
## $ smoothness_se           <dbl> 0.006399, 0.005225, 0.006150, 0.009110, 0.011490, 0.0...
## $ compactness_se          <dbl> 0.049040, 0.013080, 0.040060, 0.074580, 0.024610, 0.0...
## $ concavity_se            <dbl> 0.05373, 0.01860, 0.03832, 0.05661, 0.05688, 0.03672,...
## $ concave.points_se       <dbl> 0.015870, 0.013400, 0.020580, 0.018670, 0.018850, 0.0...
## $ symmetry_se             <dbl> 0.03003, 0.01389, 0.02250, 0.05963, 0.01756, 0.02165,...
## $ fractal_dimension_se    <dbl> 0.006193, 0.003532, 0.004571, 0.009208, 0.005115, 0.0...
## $ radius_worst            <dbl> 25.38, 24.99, 23.57, 14.91, 22.54, 15.47, 22.88, 17.0...
## $ texture_worst           <dbl> 17.33, 23.41, 25.53, 26.50, 16.67, 23.75, 27.66, 28.1...
## $ perimeter_worst         <dbl> 184.60, 158.80, 152.50, 98.87, 152.20, 103.40, 153.20...
## $ area_worst              <dbl> 2019.0, 1956.0, 1709.0, 567.7, 1575.0, 741.6, 1606.0,...
## $ smoothness_worst        <dbl> 0.1622, 0.1238, 0.1444, 0.2098, 0.1374, 0.1791, 0.144...
## $ compactness_worst       <dbl> 0.6656, 0.1866, 0.4245, 0.8663, 0.2050, 0.5249, 0.257...
## $ concavity_worst         <dbl> 0.71190, 0.24160, 0.45040, 0.68690, 0.40000, 0.53550,...
## $ concave.points_worst    <dbl> 0.26540, 0.18600, 0.24300, 0.25750, 0.16250, 0.17410,...
## $ symmetry_worst          <dbl> 0.4601, 0.2750, 0.3613, 0.6638, 0.2364, 0.3985, 0.306...
## $ fractal_dimension_worst <dbl> 0.11890, 0.08902, 0.08758, 0.17300, 0.07678, 0.12440,...
```

```r
# Replace the dot sign in column names with underscore so that all column names are in uniform form.
names(fna_new) <- str_replace_all(names(fna_new), '[.]', '_')

# Conver the response variable 'diagnosis' into factor.
fna_new$diagnosis <- as.factor(fna_new$diagnosis)

glimpse(fna_new)
```

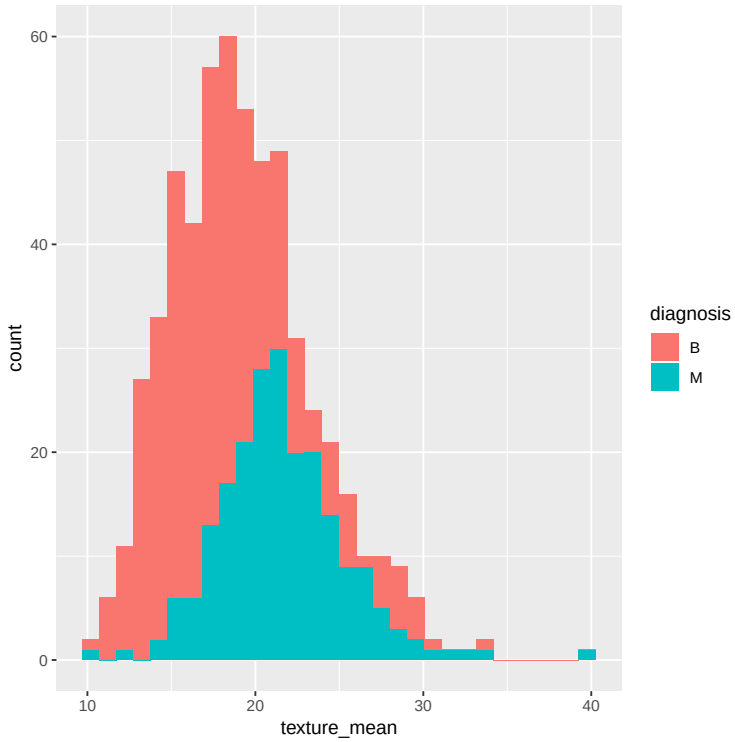
```
## Observations: 569
## Variables: 31
## $ diagnosis               <fct> M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M, M,...
## $ radius_mean             <dbl> 17.990, 20.570, 19.690, 11.420, 20.290, 12.450, 18.25...
## $ texture_mean            <dbl> 10.38, 17.77, 21.25, 20.38, 14.34, 15.70, 19.98, 20.8...
## $ perimeter_mean          <dbl> 122.80, 132.90, 130.00, 77.58, 135.10, 82.57, 119.60,...
## $ area_mean               <dbl> 1001.0, 1326.0, 1203.0, 386.1, 1297.0, 477.1, 1040.0,...
## $ smoothness_mean         <dbl> 0.11840, 0.08474, 0.10960, 0.14250, 0.10030, 0.12780,...
## $ compactness_mean        <dbl> 0.27760, 0.07864, 0.15990, 0.28390, 0.13280, 0.17000,...
## $ concavity_mean          <dbl> 0.30010, 0.08690, 0.19740, 0.24140, 0.19800, 0.15780,...
## $ concave_points_mean     <dbl> 0.14710, 0.07017, 0.12790, 0.10520, 0.10430, 0.08089,...
## $ symmetry_mean           <dbl> 0.2419, 0.1812, 0.2069, 0.2597, 0.1809, 0.2087, 0.179...
## $ fractal_dimension_mean  <dbl> 0.07871, 0.05667, 0.05999, 0.09744, 0.05883, 0.07613,...
## $ radius_se               <dbl> 1.0950, 0.5435, 0.7456, 0.4956, 0.7572, 0.3345, 0.446...
## $ texture_se              <dbl> 0.9053, 0.7339, 0.7869, 1.1560, 0.7813, 0.8902, 0.773...
## $ perimeter_se            <dbl> 8.589, 3.398, 4.585, 3.445, 5.438, 2.217, 3.180, 3.85...
## $ area_se                 <dbl> 153.40, 74.08, 94.03, 27.23, 94.44, 27.19, 53.91, 50....
## $ smoothness_se           <dbl> 0.006399, 0.005225, 0.006150, 0.009110, 0.011490, 0.0...
## $ compactness_se          <dbl> 0.049040, 0.013080, 0.040060, 0.074580, 0.024610, 0.0...
## $ concavity_se            <dbl> 0.05373, 0.01860, 0.03832, 0.05661, 0.05688, 0.03672,...
## $ concave_points_se       <dbl> 0.015870, 0.013400, 0.020580, 0.018670, 0.018850, 0.0...
## $ symmetry_se             <dbl> 0.03003, 0.01389, 0.02250, 0.05963, 0.01756, 0.02165,...
## $ fractal_dimension_se    <dbl> 0.006193, 0.003532, 0.004571, 0.009208, 0.005115, 0.0...
## $ radius_worst            <dbl> 25.38, 24.99, 23.57, 14.91, 22.54, 15.47, 22.88, 17.0...
## $ texture_worst           <dbl> 17.33, 23.41, 25.53, 26.50, 16.67, 23.75, 27.66, 28.1...
## $ perimeter_worst         <dbl> 184.60, 158.80, 152.50, 98.87, 152.20, 103.40, 153.20...
## $ area_worst              <dbl> 2019.0, 1956.0, 1709.0, 567.7, 1575.0, 741.6, 1606.0,...
## $ smoothness_worst        <dbl> 0.1622, 0.1238, 0.1444, 0.2098, 0.1374, 0.1791, 0.144...
## $ compactness_worst       <dbl> 0.6656, 0.1866, 0.4245, 0.8663, 0.2050, 0.5249, 0.257...
## $ concavity_worst         <dbl> 0.71190, 0.24160, 0.45040, 0.68690, 0.40000, 0.53550,...
## $ concave_points_worst    <dbl> 0.26540, 0.18600, 0.24300, 0.25750, 0.16250, 0.17410,...
## $ symmetry_worst          <dbl> 0.4601, 0.2750, 0.3613, 0.6638, 0.2364, 0.3985, 0.306...
## $ fractal_dimension_worst <dbl> 0.11890, 0.08902, 0.08758, 0.17300, 0.07678, 0.12440,...
```

```r
characteristics <- names(fna_new)[-1]
for (i in 1:30){
  p <- ggplot(fna_new, aes(x = get(characteristics[i]), fill = diagnosis)) +
    geom_histogram() +
    xlab(characteristics[i])
  print(p)
}
```

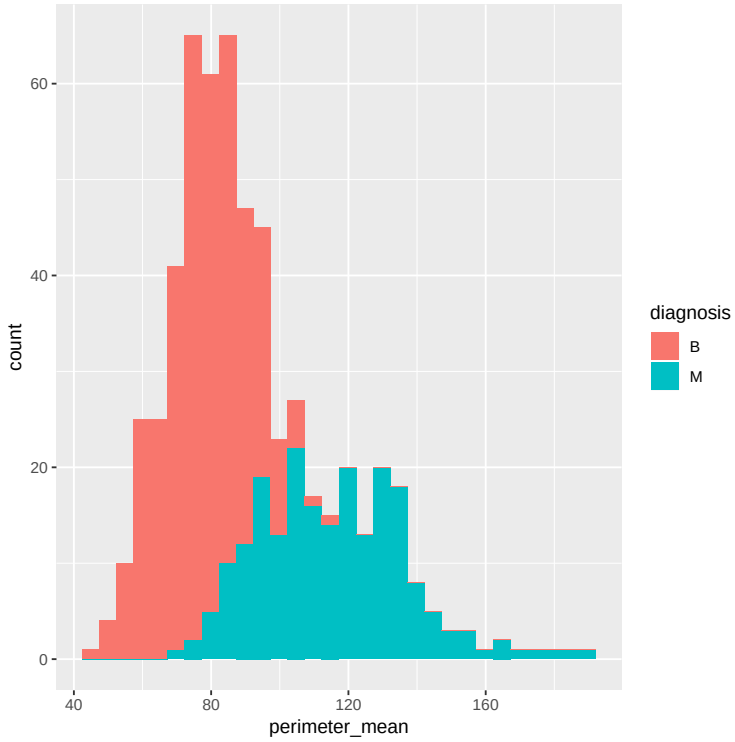
```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



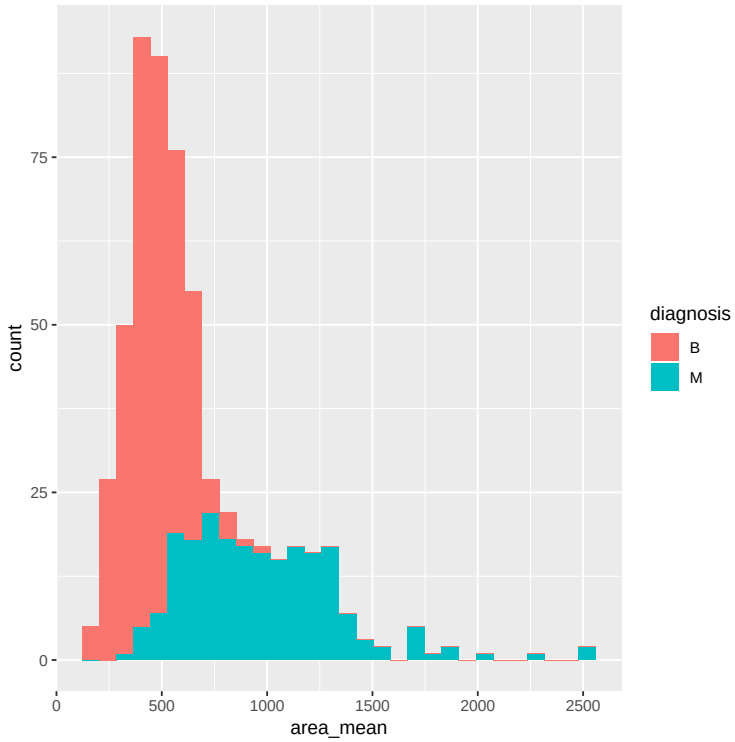

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



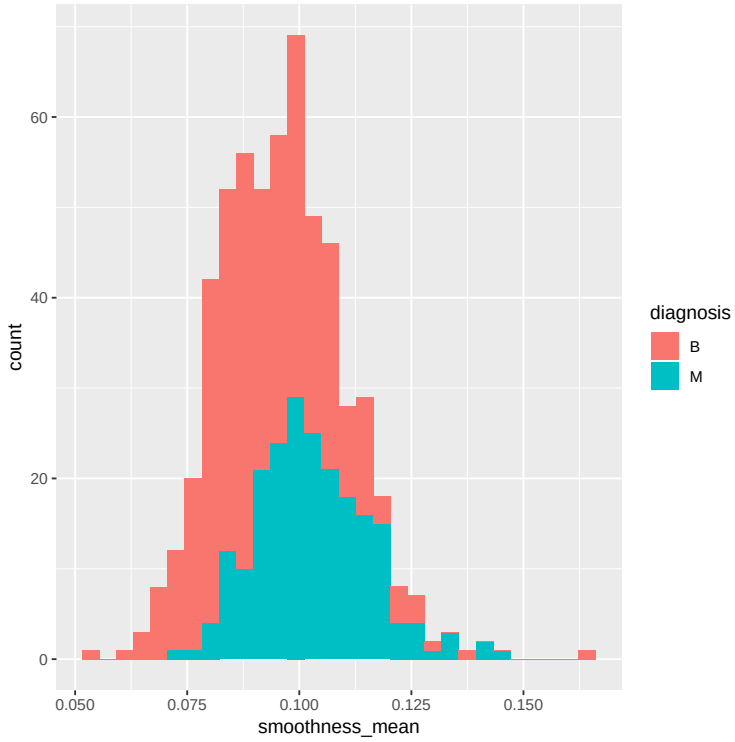

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



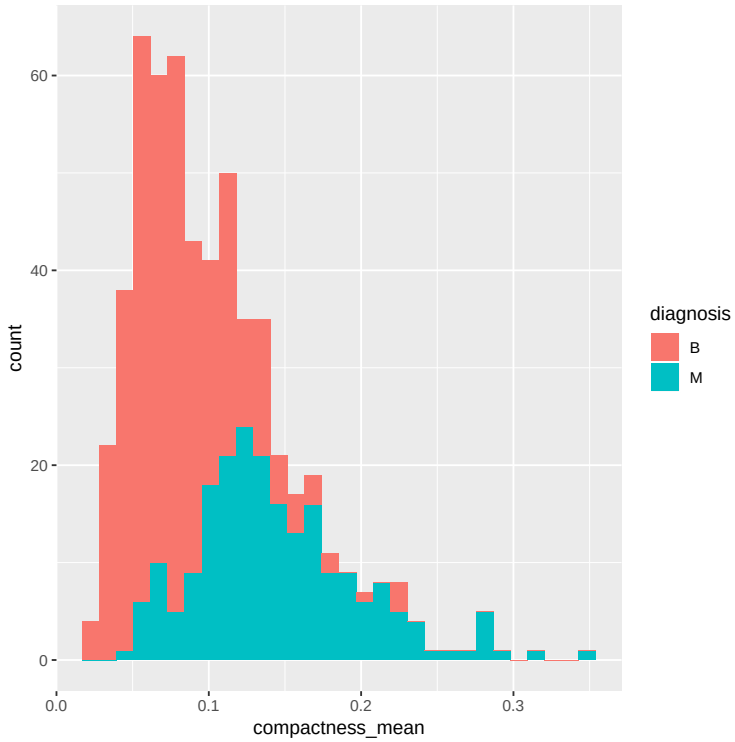

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



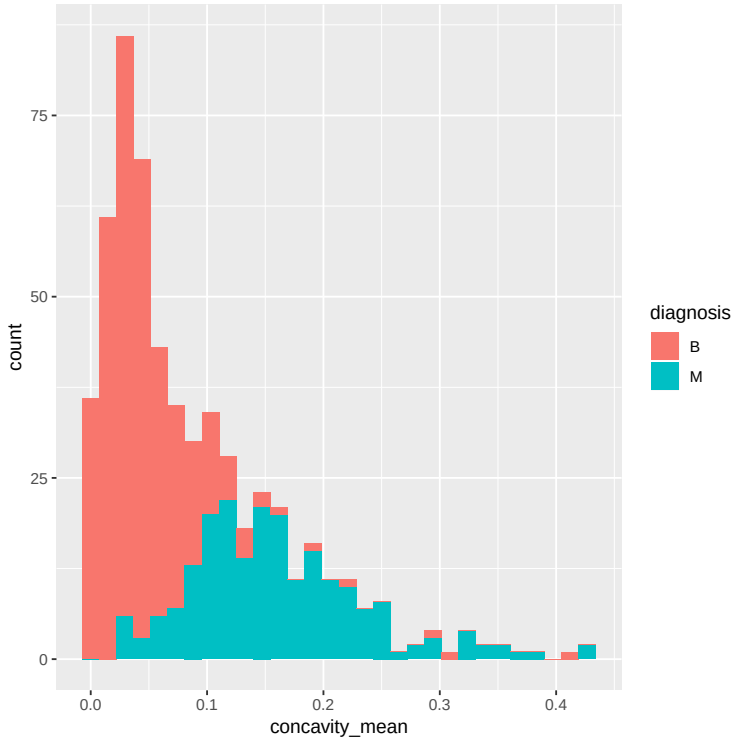

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



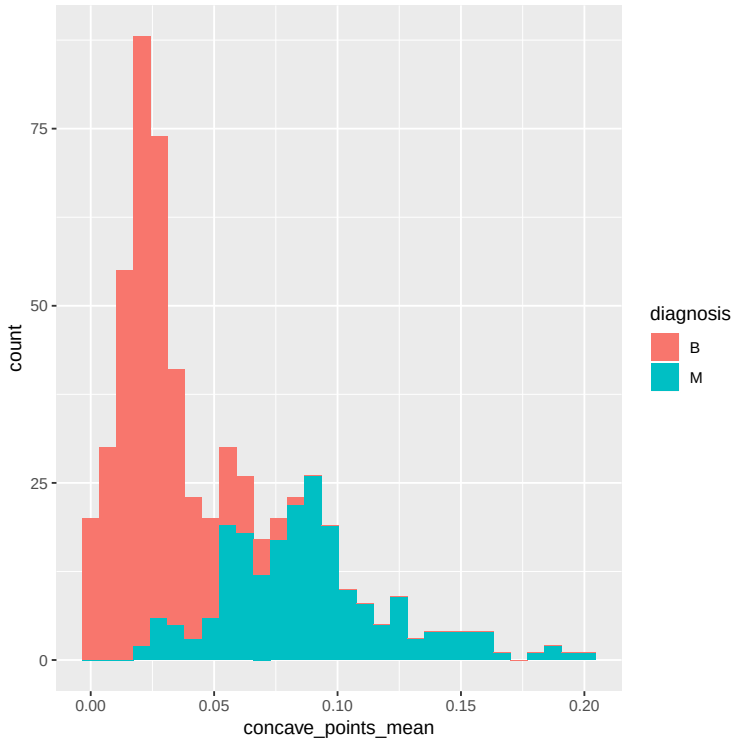

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



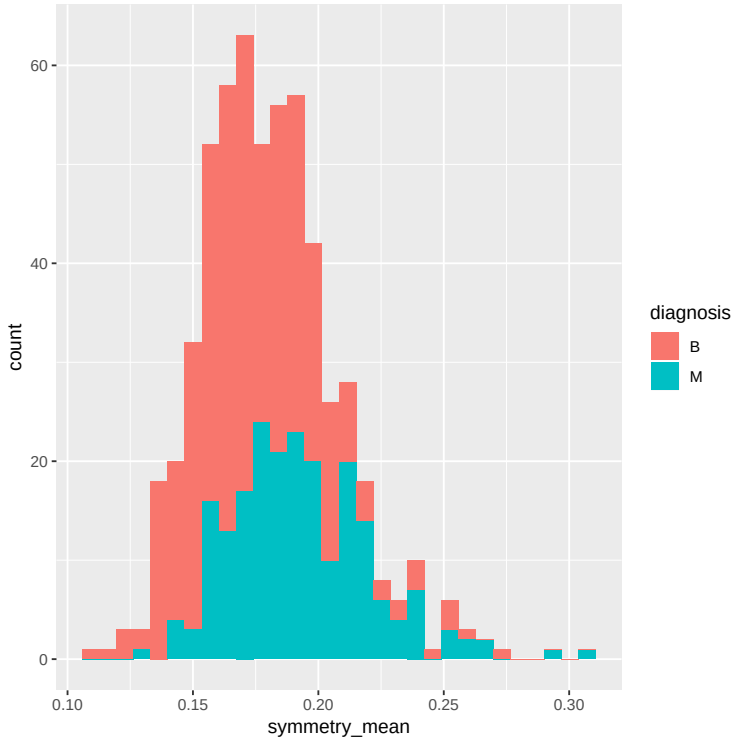

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



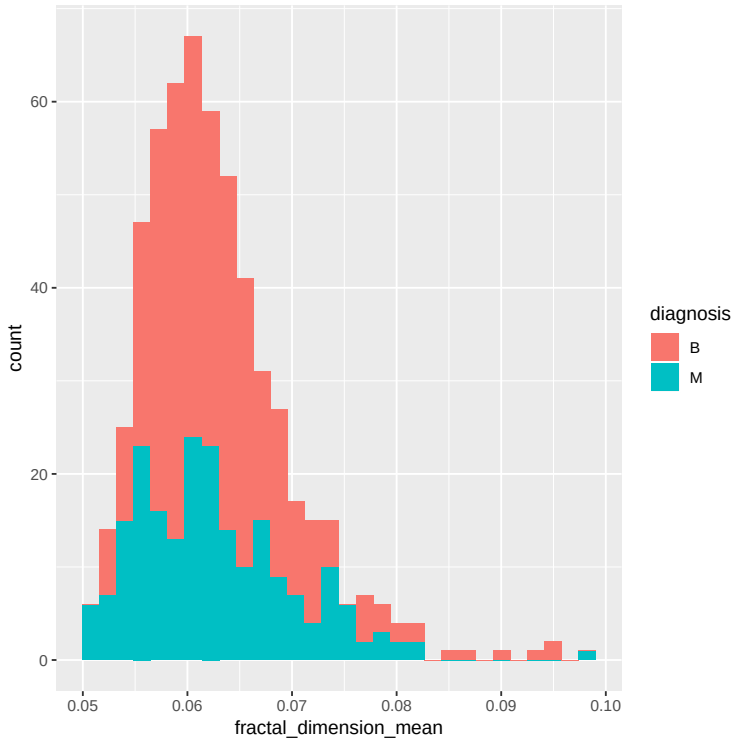

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



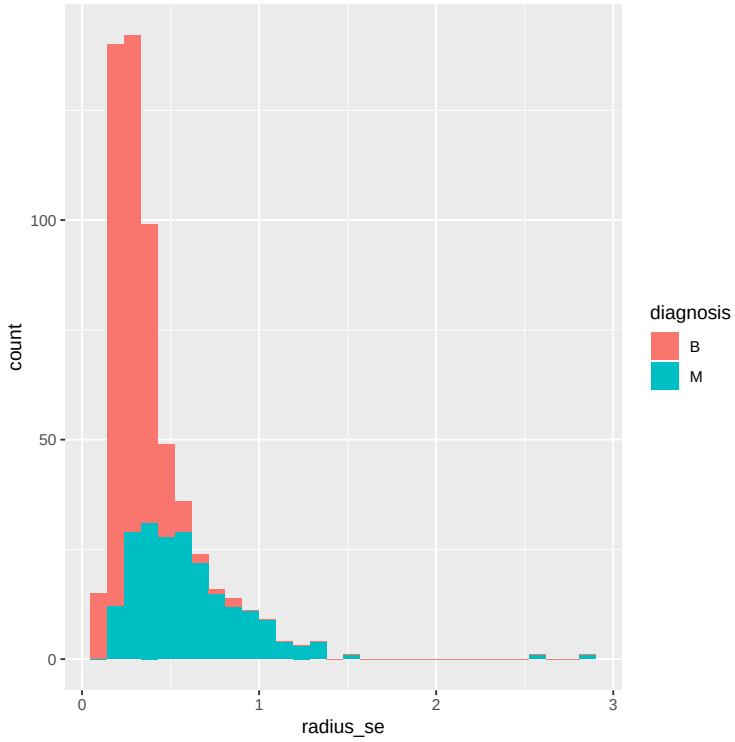

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



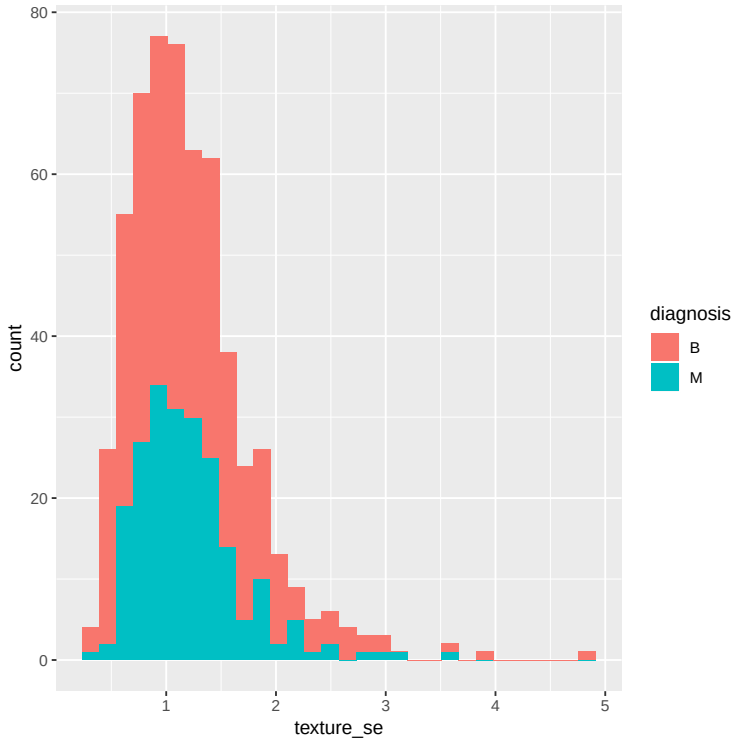

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



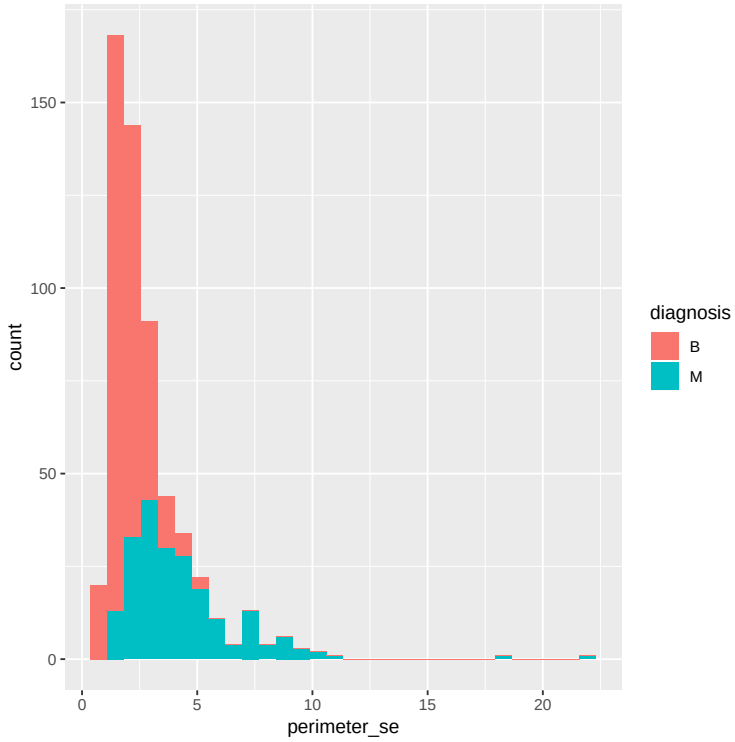

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



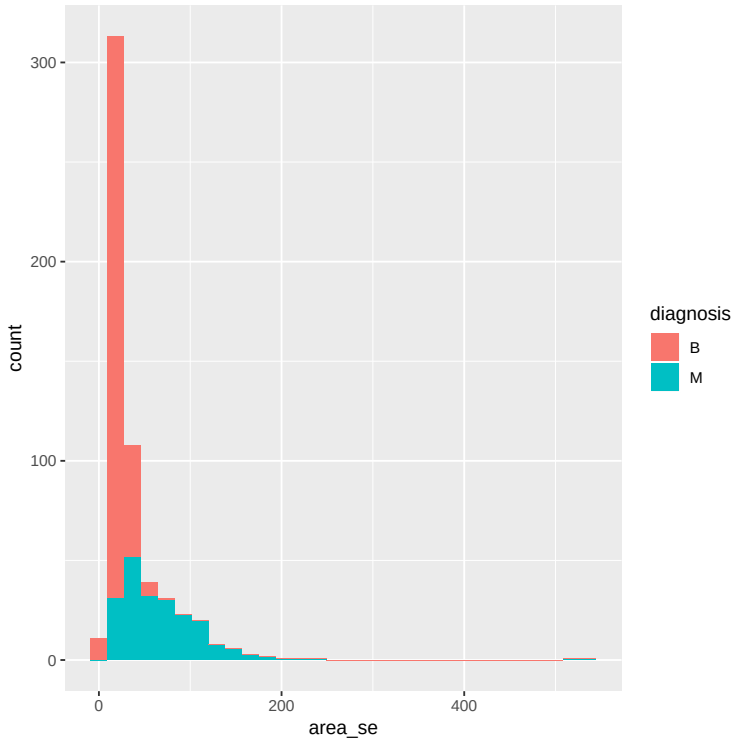

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



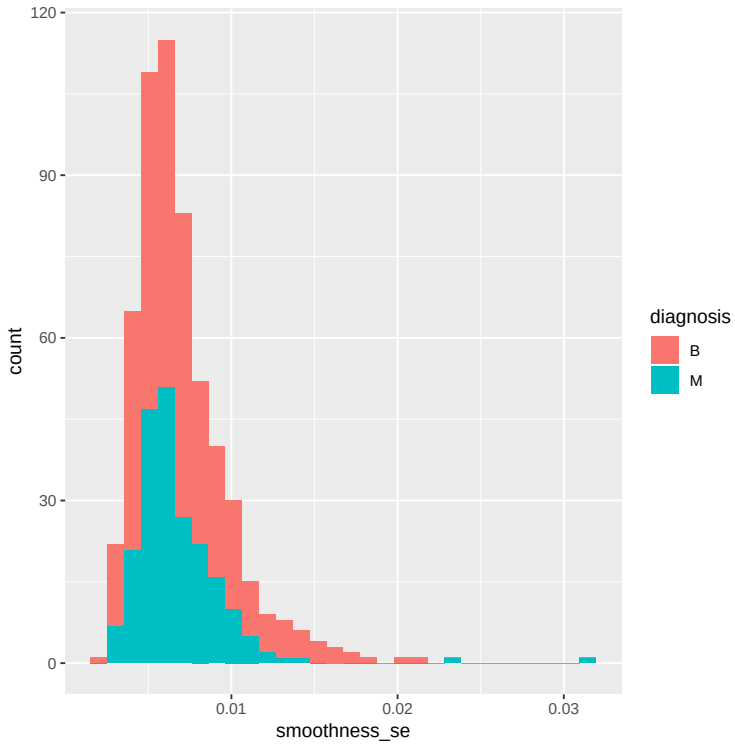

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



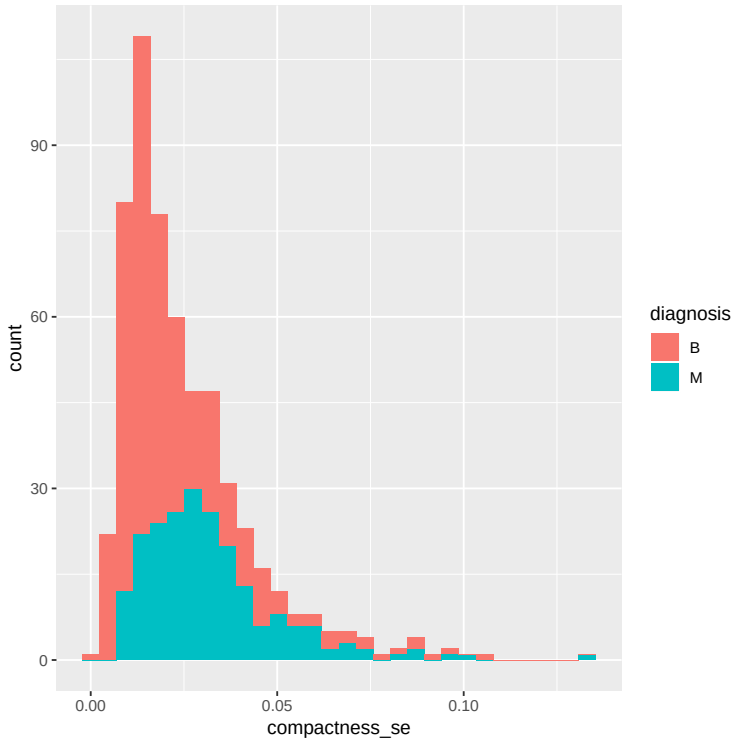

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



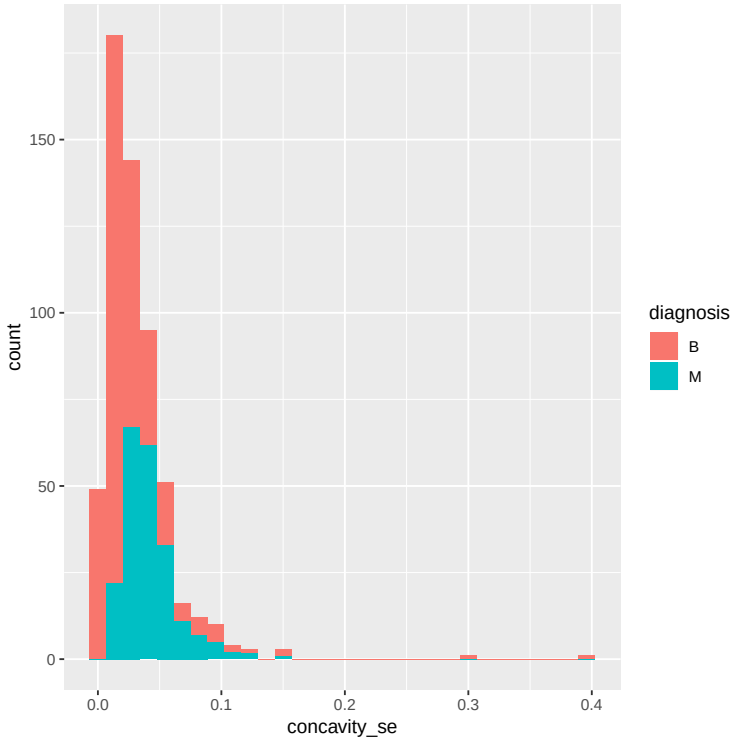

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



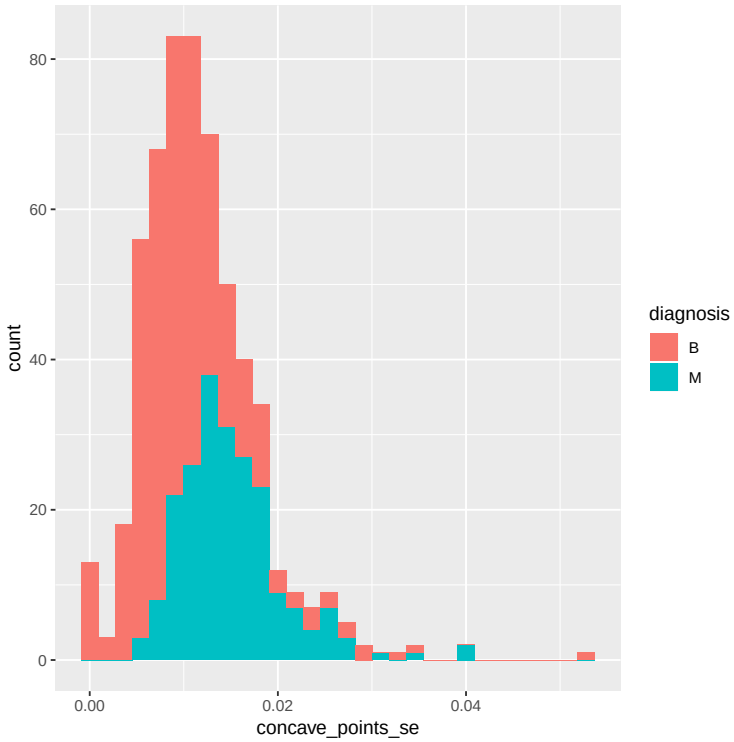

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



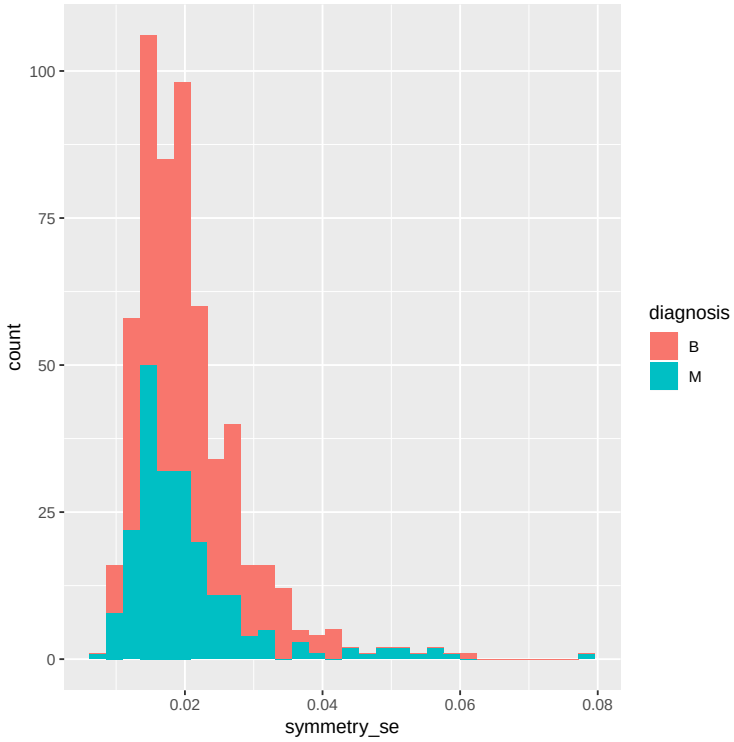

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



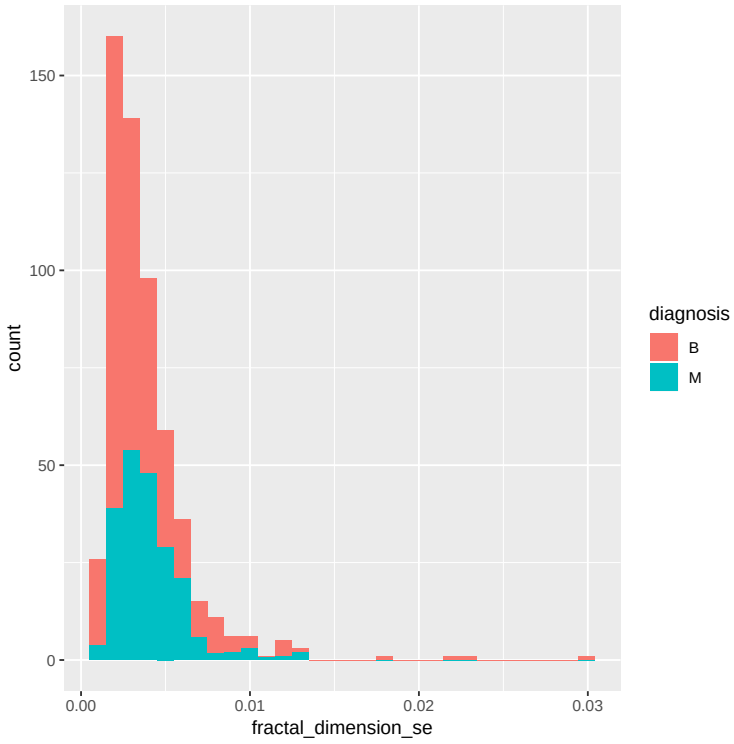

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



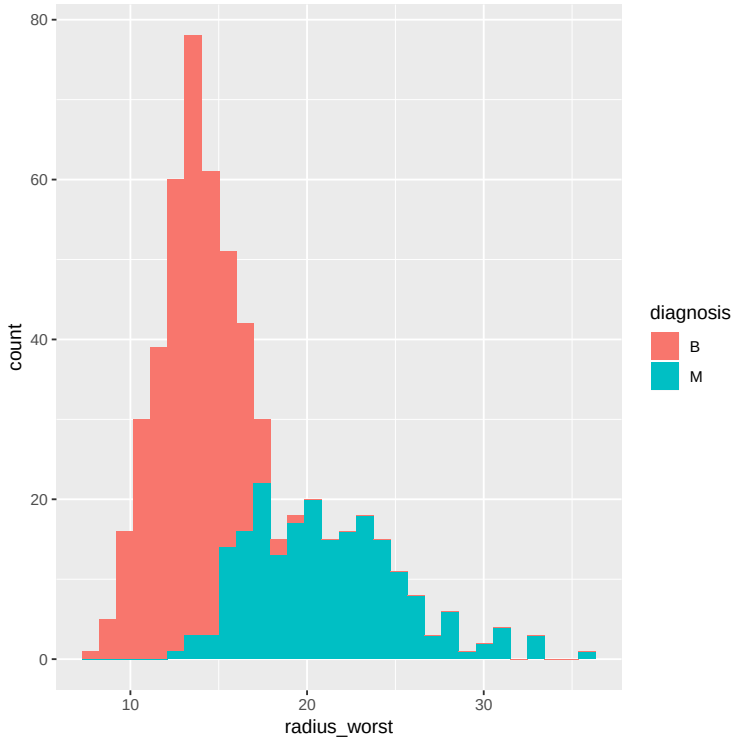

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



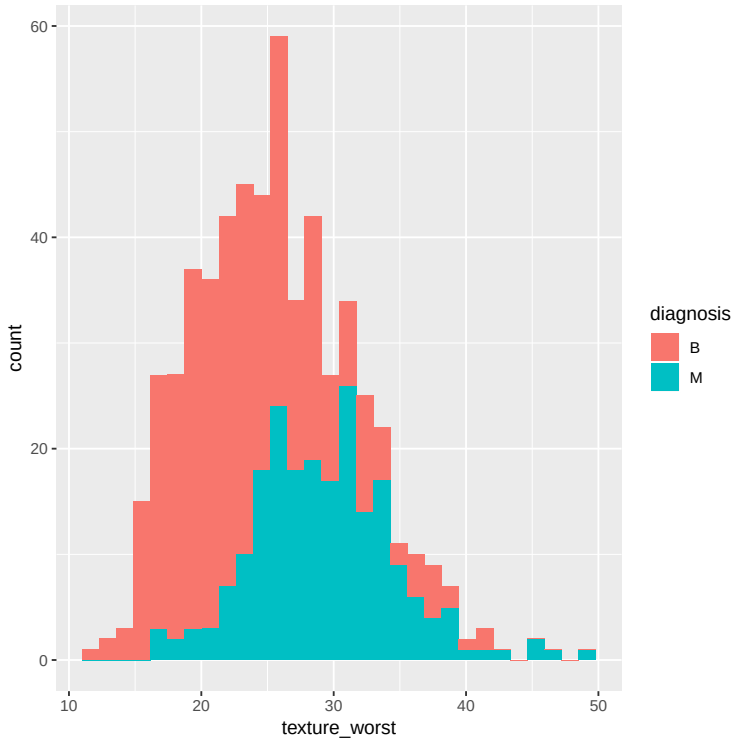

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



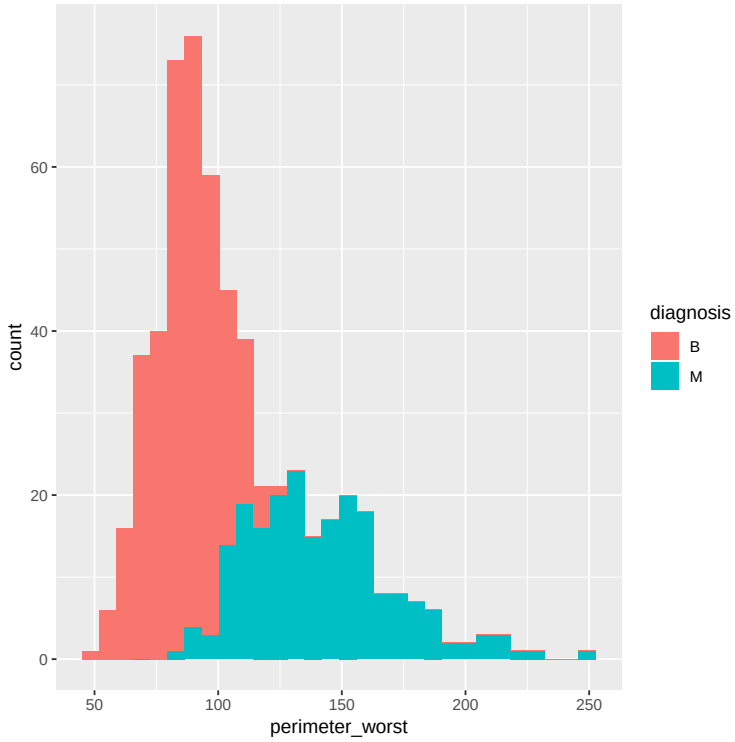

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



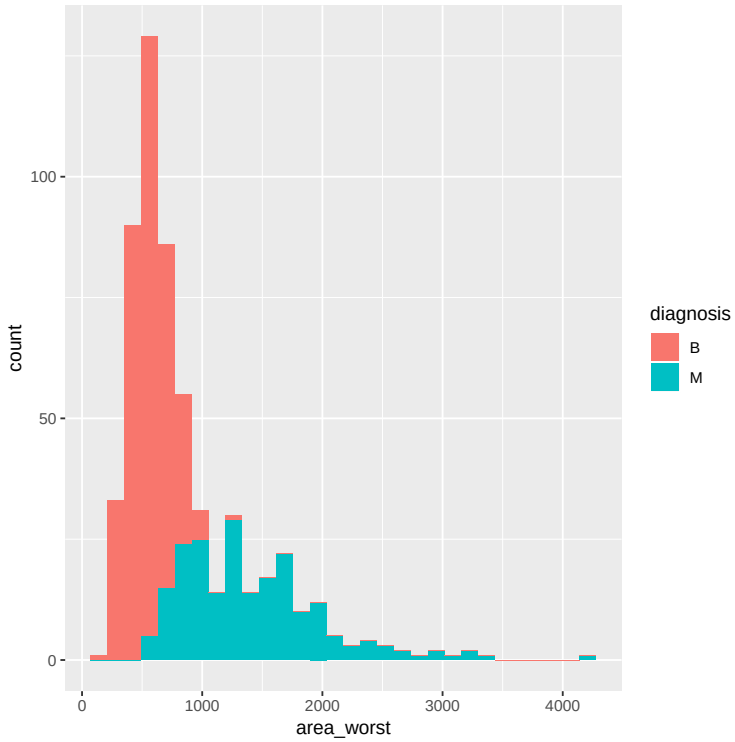

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



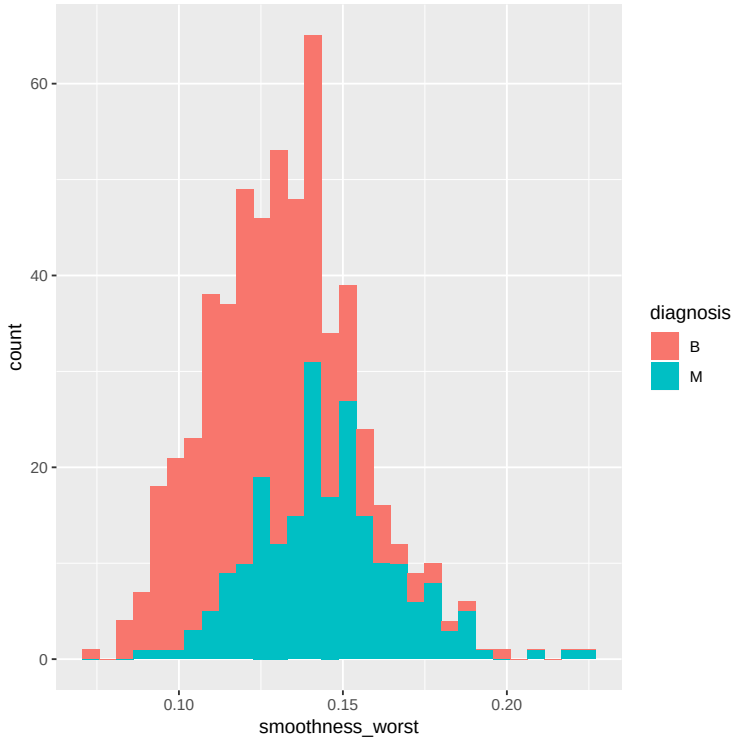

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



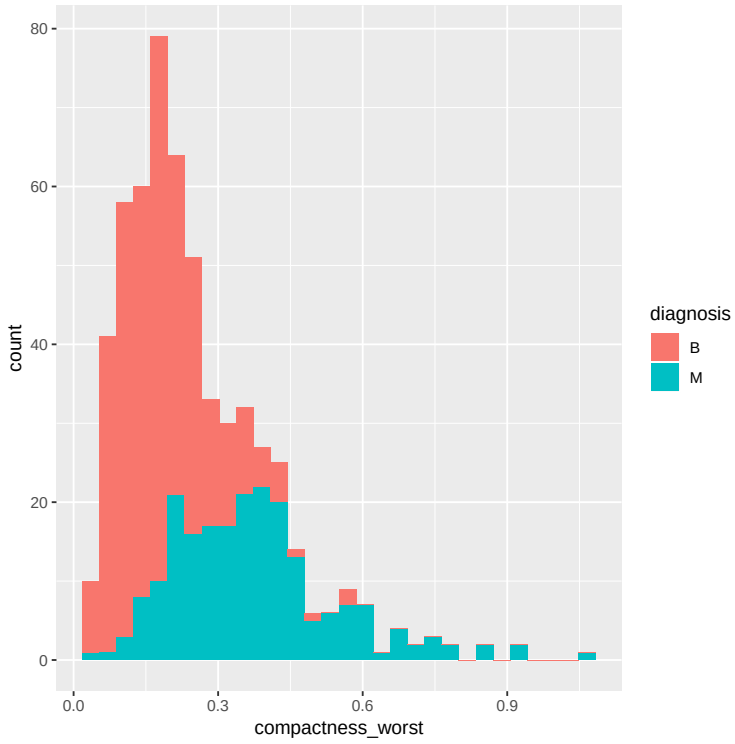

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



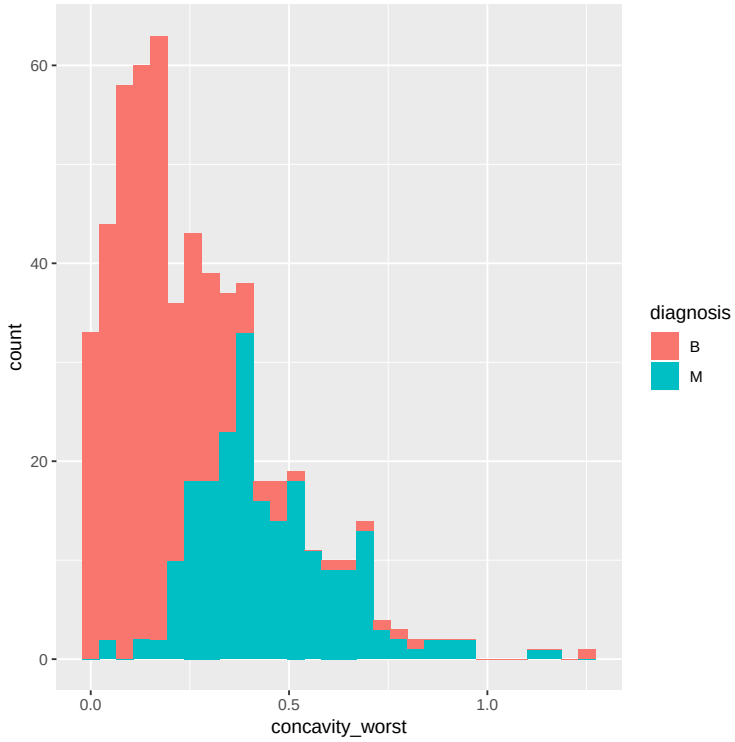

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



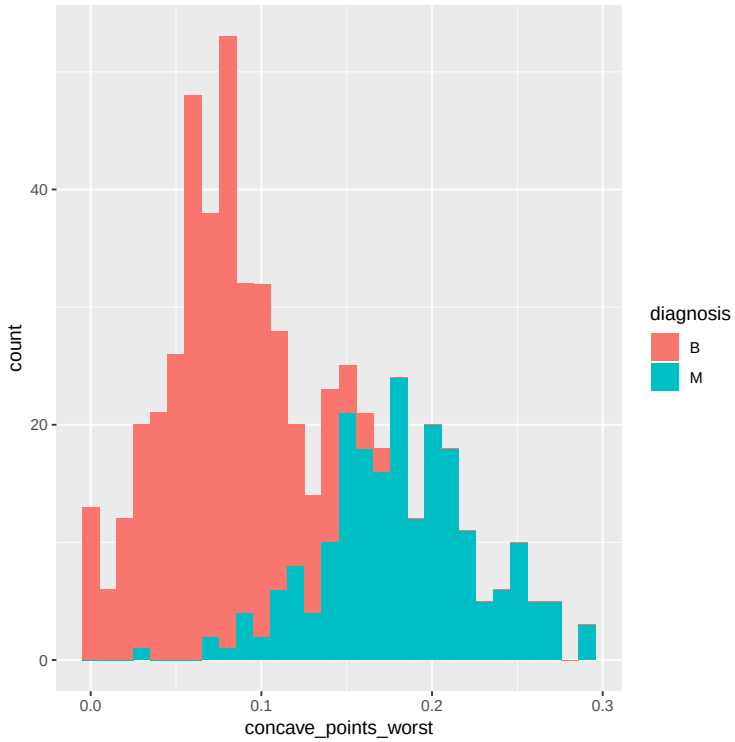

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



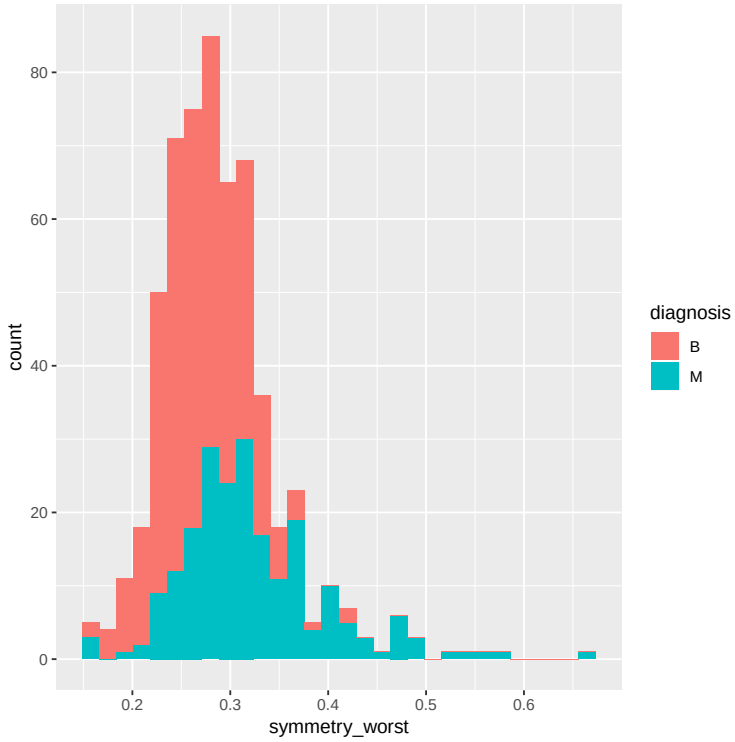

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



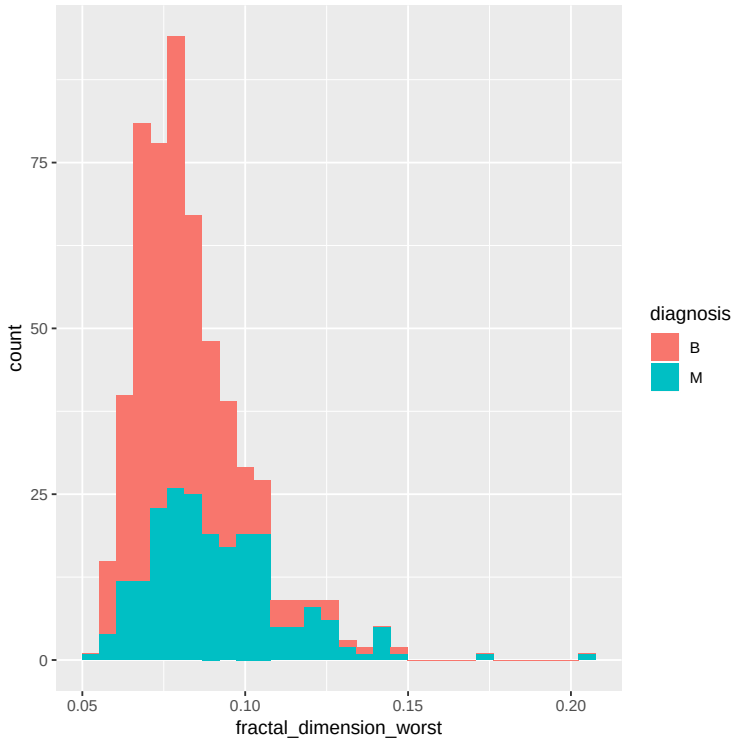

```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```




```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```




```
## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



The R session information (including the OS info, R version and all packages used):

```
sessionInfo()
Sys.time()
```