Data 607 Subsetting Datasets Lab

Mary Anna Kivenson

2/3/2019

Importing the Bridges Dataset

Lab Info: file:///Users/MaryAnna/Downloads/Hands%20On%20Lab%20Subsetting%20Datasets(1)%20(3).pdf

Data Source: https://archive.ics.uci.edu/ml/datasets/Pittsburgh+Bridges

```
bridges <-
  read.csv("https://archive.ics.uci.edu/ml/machine-learning-databases/bridges/bridges.data.version1",
           header= FALSE, sep=",")
```

Renaming the Columns

```
colnames(bridges) <- c("IDENTIF","RIVER","LOCATION","ERECTED","PURPOSE","LENGTH","LANES",
                       "CLEAR-G","T-OR-D","MATERIAL","SPAN","REL-L","TYPE")
head(bridges)
```

```
##   IDENTIF RIVER LOCATION ERECTED  PURPOSE LENGTH LANES CLEAR-G  T-OR-D
## 1      E1     M        3    1818  HIGHWAY      ?     2       N THROUGH
## 2      E2     A       25    1819  HIGHWAY   1037     2       N THROUGH
## 3      E3     A       39    1829 AQUEDUCT      ?     1       N THROUGH
## 4      E5     A       29    1837  HIGHWAY   1000     2       N THROUGH
## 5      E6     M       23    1838  HIGHWAY      ?     2       N THROUGH
## 6      E7     A       27    1840  HIGHWAY    990     2       N THROUGH
##   MATERIAL   SPAN REL-L TYPE
## 1     WOOD  SHORT     S WOOD
## 2     WOOD  SHORT     S WOOD
## 3     WOOD      ?     S WOOD
## 4     WOOD  SHORT     S WOOD
## 5     WOOD      ?     S WOOD
## 6     WOOD MEDIUM     S WOOD
```

Subsetting the Data

```
bridges_subset <- subset(bridges, PURPOSE=="HIGHWAY", select=c(ERECTED,PURPOSE,MATERIAL))
bridges_subset
```

```
##     ERECTED PURPOSE MATERIAL
## 1      1818 HIGHWAY     WOOD
## 2      1819 HIGHWAY     WOOD
## 4      1837 HIGHWAY     WOOD
## 5      1838 HIGHWAY     WOOD
## 6      1840 HIGHWAY     WOOD
## 8      1846 HIGHWAY     IRON
## 10     1851 HIGHWAY     WOOD
```

```
## 12     1856 HIGHWAY     WOOD
## 13     1856 HIGHWAY     WOOD
## 15     1859 HIGHWAY     IRON
## 18     1866 HIGHWAY     WOOD
## 19     1870 HIGHWAY     WOOD
## 21     1876 HIGHWAY    STEEL
## 22     1876 HIGHWAY     WOOD
## 28     1884 HIGHWAY    STEEL
## 29     1884 HIGHWAY    STEEL
## 30     1887 HIGHWAY     IRON
## 33     1889 HIGHWAY     IRON
## 34     1890 HIGHWAY     IRON
## 35     1890 HIGHWAY    STEEL
## 36     1891 HIGHWAY     IRON
## 38     1892 HIGHWAY    STEEL
## 40     1893 HIGHWAY    STEEL
## 41     1894 HIGHWAY     IRON
## 42     1895 HIGHWAY    STEEL
## 43     1896 HIGHWAY    STEEL
## 44     1896 HIGHWAY    STEEL
## 48     1900 HIGHWAY    STEEL
## 49     1900 HIGHWAY    STEEL
## 51     1902 HIGHWAY    STEEL
## 59     1908 HIGHWAY    STEEL
## 60     1909 HIGHWAY    STEEL
## 61     1909 HIGHWAY    STEEL
## 63     1911 HIGHWAY    STEEL
## 67     1915 HIGHWAY    STEEL
## 71     1923 HIGHWAY    STEEL
## 72     1924 HIGHWAY    STEEL
## 73     1926 HIGHWAY    STEEL
## 74     1926 HIGHWAY    STEEL
## 75     1927 HIGHWAY    STEEL
## 76     1927 HIGHWAY    STEEL
## 77     1927 HIGHWAY    STEEL
## 78     1927 HIGHWAY    STEEL
## 79     1928 HIGHWAY    STEEL
## 80     1928 HIGHWAY    STEEL
## 81     1928 HIGHWAY    STEEL
## 82     1928 HIGHWAY    STEEL
## 83     1931 HIGHWAY    STEEL
## 84     1931 HIGHWAY    STEEL
## 85     1931 HIGHWAY    STEEL
## 86     1937 HIGHWAY    STEEL
## 87     1939 HIGHWAY    STEEL
## 88     1945 HIGHWAY    STEEL
## 90     1945 HIGHWAY    STEEL
## 91     1945 HIGHWAY    STEEL
## 92     1945 HIGHWAY    STEEL
## 94     1950 HIGHWAY    STEEL
## 95     1951 HIGHWAY    STEEL
## 96     1951 HIGHWAY    STEEL
## 97     1951 HIGHWAY    STEEL
## 98     1955 HIGHWAY    STEEL
```

```
## 99     1955 HIGHWAY    STEEL
## 100    1959 HIGHWAY    STEEL
## 101    1959 HIGHWAY    STEEL
## 102    1961 HIGHWAY    STEEL
## 103    1962 HIGHWAY    STEEL
## 104    1969 HIGHWAY    STEEL
## 105    1975 HIGHWAY    STEEL
## 106    1978 HIGHWAY    STEEL
## 107    1982 HIGHWAY        ?
## 108    1986 HIGHWAY        ?
```

Summarizing the Data

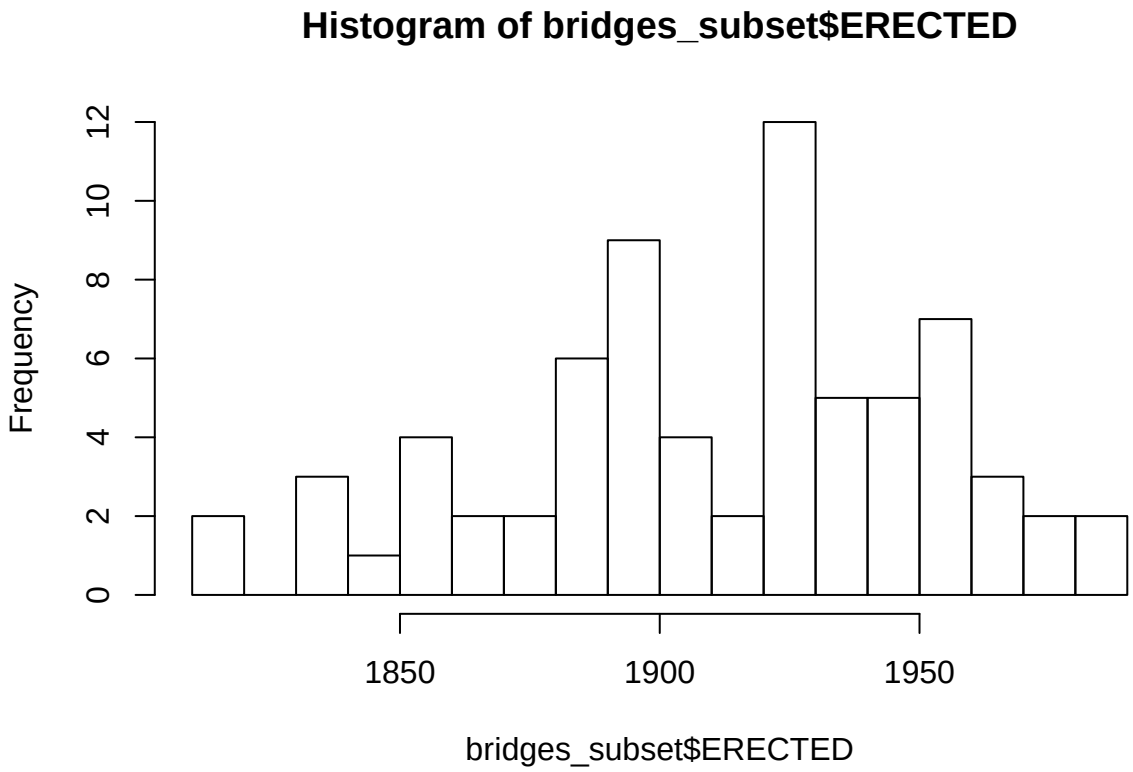
```
summary(bridges_subset)
```

```
##     ERECTED         PURPOSE    MATERIAL
##  Min.   :1818   AQUEDUCT: 0   ?    : 2
##  1st Qu.:1890   HIGHWAY :71   IRON : 7
##  Median :1923   RR      : 0   STEEL:51
##  Mean   :1912   WALK    : 0   WOOD :11
##  3rd Qu.:1945
##  Max.   :1986
```

Histogram of Erected Bridges - Timeline

This histogram shows the amount of highways built in each grouping of 10 years. Overall, the distribution of highways built over the years is left skewed - the majority of bridges were built prior to 1930. The historgram is also bimodal, showing bursts in highway construction between 1890 - 1900 and 1920 - 1930.

```
hist(bridges_subset$ERECTED, breaks = 20)
```



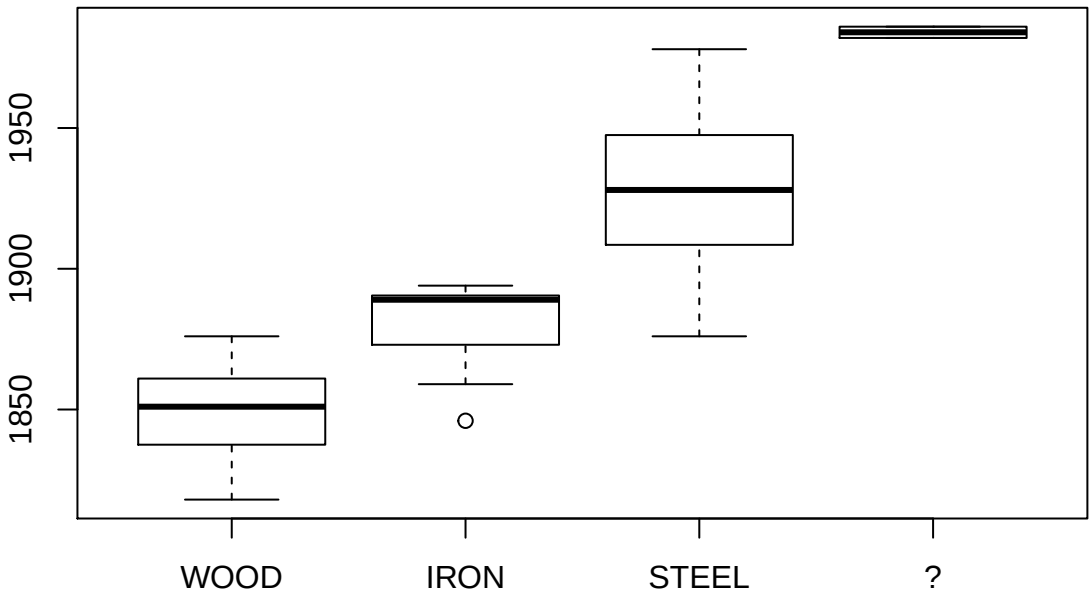
Year vs Highway Material Boxplot

The boxplot below shows the median years that each highway material was used. Based on the plot, it seems like the prefered highway material was wood until 1875, followed by iron until 1900 and steel from 1900 onwards.

```
require(MASS)
```

```
## Loading required package: MASS
```

```
data(iris)
boxplot(bridges_subset$ERECTED ~ bridges_subset$MATERIAL, at=rank(tapply(bridges_subset$ERECTED, bridges
```


1850
1900
1950
WOOD
IRON
STEEL
?