# 005-beamer-presentations
## Forced Expiratory Volume and Smoking

Sahir Rai Bhatnagar[1]

May 28, 2014

---

[1] https://github.com/sahirbhatnagar/knitr-tutorial

# Abstract

## Forced Expiratory Volume and Smoking

Presenting research is an important part of a statisticians life. We illustrate the use of Beamer presentations and `knitr` (Xie, 2015, 2013, 2014) using data from a study that aimed to assess the relationship between subjects forced expiratory volume (FEV) and their current smoking status. In this problem the measured outcome of interest is forced expiratory volume (FEV), which is, essentially, the amount of air an individual can exhale in the first second of a forceful breath. The data recorded in the dataset include the following: FEV (liters), AGE (years), HEIGHT (inches), GENDER (M/F), SMOKE (Y/N) (Kahn, 2005).

# Fivenumber Summary of Sex-Education Combinations

A very powerful way of getting custom summary information by multiple categories is via the `plyr` package (Wickham, 2011).

# Fivenumber Summary of Sex-Education Combinations

A very powerful way of getting custom summary information by multiple categories is via the `plyr` package (Wickham, 2011).

It allows you to subset the data and perform the operations in a single step

# Fivenumber Summary of Sex-Education Combinations

A very powerful way of getting custom summary information by multiple categories is via the `plyr` package (Wickham, 2011).

It allows you to subset the data and perform the operations in a single step

```
fev <- read.csv("lung.csv")

fev$edu <- cut(fev$age, breaks = c(2, 5, 10, 13,
    Inf), labels = c("preschool", "primary", "middle",
    "highschool"))

ddply(fev, .(edu, sex), summarise, min = min(fev),
    `1st` = quantile(fev, 1/4), median = median(fev),
    mean = mean(fev), `3rd` = quantile(fev, 3/4),
    max = max(fev))
```

# Fivenumber Summary of Sex-Education Combinations

```
##              edu sex   min 1st median mean 3rd max
## 1  preschool   0 0.79 1.1    1.4  1.3 1.6 1.7
## 2  preschool   1 0.80 1.5    1.8  1.6 1.8 2.1
## 3    primary   0 1.29 1.8    2.2  2.2 2.6 3.4
## 4    primary   1 1.17 1.8    2.2  2.3 2.6 4.6
## 5     middle   0 2.08 2.6    3.0  2.9 3.2 3.8
## 6     middle   1 1.69 2.9    3.4  3.5 4.1 5.2
## 7 highschool   0 2.20 2.7    3.0  3.0 3.3 3.7
## 8 highschool   1 2.28 3.7    4.2  4.2 4.5 5.8
```

# The Power of R Graphics

A very powerful graphics package in R is ggplot2 (Wickham, 2009).

# The Power of R Graphics

A very powerful graphics package in R is ggplot2 (Wickham, 2009).

Similar in spirit to the plyr package, subsetting and plotting are done simultaneously.

# The Power of R Graphics

A very powerful graphics package in R is ggplot2 (Wickham, 2009).

Similar in spirit to the plyr package, subsetting and plotting are done simultaneously.

See http://docs.ggplot2.org/current/ for documentation

http://www.cookbook-r.com/Graphs/ is also a very good resource with examples

# Boxplots of Sex-Education Combinations

```r
# change 0/1 to male / female
fev$gender <- sapply(fev$sex, function(i) if (i ==
    1) "Male" else "Female")

# check that edu and gender are
# Binary/Factor/Character variables
str(fev)

## 'data.frame': 654 obs. of  7 variables:
##  $ age   : int  9 8 7 9 9 8 6 6 8 9 ...
##  $ fev   : num  1.71 1.72 1.72 1.56 1.9 ...
##  $ height: num  57 67.5 54.5 53 57 61 58 56 58.5 60 ...
##  $ sex   : int  0 0 0 1 1 0 0 0 0 0 ...
##  $ smoke : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ edu   : Factor w/ 4 levels "preschool","primary",..:
##  $ gender: chr  "Female" "Female" "Female" "Male" ...
```
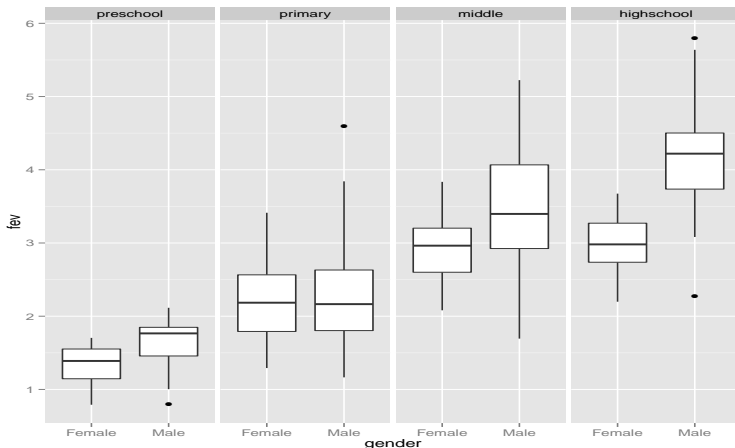
# Boxplots of Sex-Education Combinations

```
p <- ggplot(fev, aes(x = gender, y = fev)) + geom_boxplot()
p + facet_grid(~edu)
```
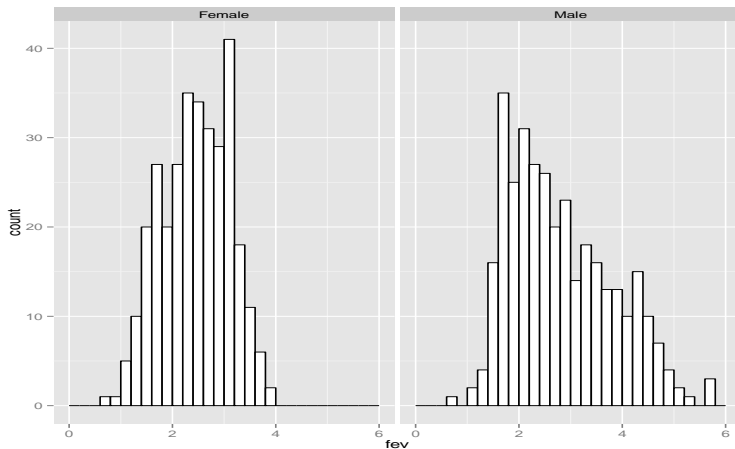
# Histograms by Gender

```r
# initiate ggplot, specify breaks
m <- ggplot(fev, aes(x = fev)) + geom_histogram(colour = "black",
    fill = "white", breaks = seq(0, 6, 0.2))
```

# Histograms by Gender

```r
# initiate ggplot, specify breaks
m <- ggplot(fev, aes(x = fev)) + geom_histogram(colour = "black",
    fill = "white", breaks = seq(0, 6, 0.2))
```

```r
# plot FEV by gender
m + facet_grid(~gender)
```
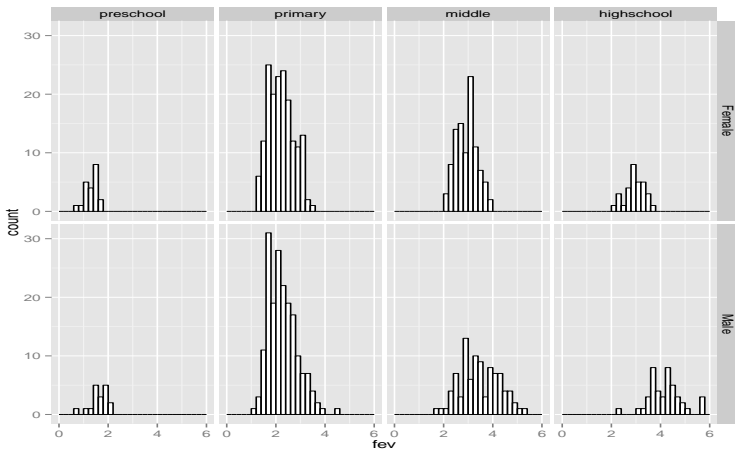
# Histograms by Gender

```
m <- ggplot(fev, aes(x = fev)) + geom_histogram(colour = "black",
    fill = "white", breaks = seq(0, 6, 0.2))
m + facet_grid(~gender)
```

# Histograms by Gender-Education Combinations

```
# where 'm' is the same as previous slide
m + facet_grid(gender ~ edu)
```
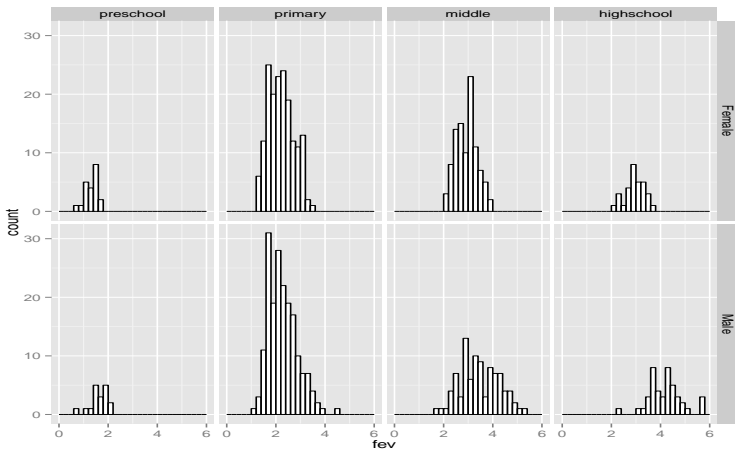
# Histograms by Gender-Education Combinations

```
# where 'm' is the same as previous slide
m + facet_grid(gender ~ edu)
```



Question: *What is the problem with this plot?*

# Table of Gender-Education Combinations Counts

```
xtabs(~edu + gender, data = fev)

##               gender
## edu           Female Male
##   preschool       21   18
##   primary        168  183
##   middle          98   92
##   highschool      31   43
```

# Density of Gender-Education Combinations

To make the plots more comparable plot their densities

005-beamer-
presentations

Introduction
 Abstract

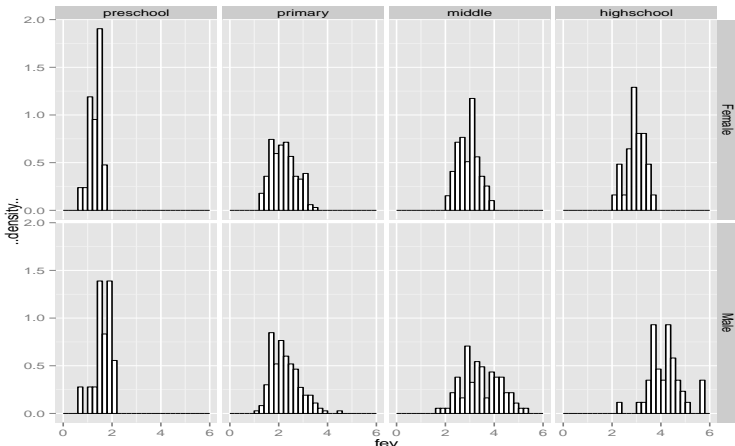Fivenumber
Summary

Boxplots

Histograms

References

# Density of Gender-Education Combinations

To make the plots more comparable plot their densities

```
# where 'm' is the same as previous slides
m + aes(y = ..density..) + facet_grid(gender ~ edu)
```

# Histogram vs. Density of FEV

# References

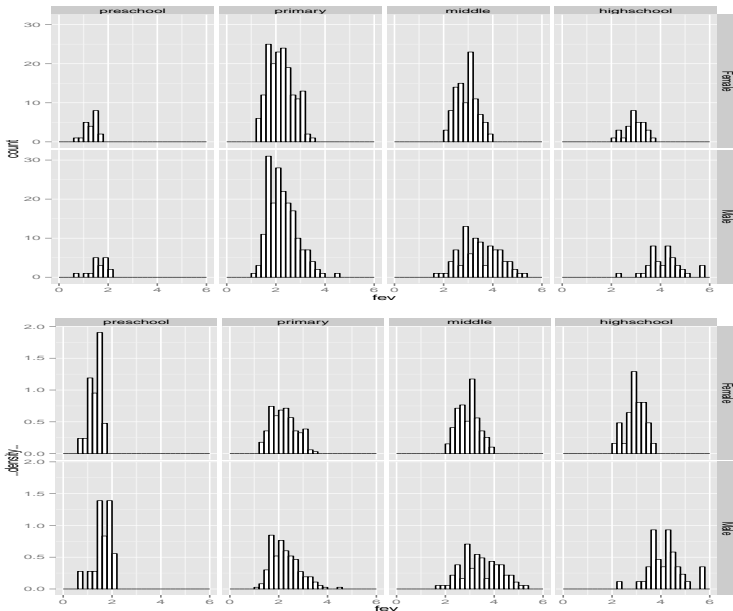Michael Kahn. An exhalent problem for teaching statistics. *The Journal of Statistical Education*, 13(2), 2005.

Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. URL `http://had.co.nz/ggplot2/book`.

Hadley Wickham. The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29, 2011. URL `http://www.jstatsoft.org/v40/i01/`.

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2013. URL `http://yihui.name/knitr/`. ISBN 978-1482203530.

Yihui Xie. knitr: A comprehensive tool for reproducible research in R. In Victoria Stodden, Friedrich Leisch, and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2014. URL `http://www.crcpress.com/product/isbn/9781466561595`. ISBN 978-1466561595.

Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2015. URL `http://yihui.name/knitr/`. R package version 1.10.5.

# Session Info

```r
print(sessionInfo(), locale = FALSE)

## R version 3.2.0 (2015-04-16)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 14.04.2 LTS
##
## attached base packages:
## [1] methods    stats      graphics  grDevices
## [5] utils      datasets   base
##
## other attached packages:
## [1] ggplot2_1.0.1 plyr_1.8.2    knitr_1.10.5
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.11.6       digest_0.6.8
##  [3] MASS_7.3-39       grid_3.2.0
##  [5] gtable_0.1.2      formatR_1.2
##  [7] magrittr_1.5      evaluate_0.7
##  [9] scales_0.2.4      highr_0.5
## [11] stringi_0.4-1     reshape2_1.4.1
## [13] labeling_0.3      proto_0.3-10
## [15] tools_3.2.0       stringr_1.0.0
## [17] munsell_0.4.2     colorspace_1.2-6
```