Object-Oriented Markdown in R to Facilitate Collaboration

Rickey Carter and Tina Gunderson

Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

6 July 2017
The Problem
Reporting Goals, Regardless of Software

1. Reproducible research
2. Facilitate team collaboration
3. Minimize programming redundancy and manual input
4. Make it easy to control and share output
SAS default output

SAS

proc ttest data=dataset;
class group;
var outcome;
run;
R t.test(outcome ~ group, data = dataset)

Welch Two Sample t-test

data:  outcome by group
t = -3.04, df = 72.52, p-value = 0.00329
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8303715  -0.1726927
sample estimates:
mean in group Group A  mean in group Group B
 -0.03783587 0.46369624
An Expanded R t-test

```r
# Get descriptive stats
dataset %>% group_by(group) %>%
  summarise(Means=mean(outcome), sd=sd(outcome),
            se=sd/sqrt(n()), min=min(outcome), max=max(outcome))

# Perform t-test
t.test(outcome ~ group, data=dataset)

# Histogram
ggplot(dataset, aes(outcome, fill=group)) +
  geom_histogram(color="black", bins=10) +
  facet_wrap(~group, ncol=1)

# Boxplot
ggplot(dataset, aes(Group, outcome)) +
  geom_boxplot() + coord_flip()
```
Some Output Options

1. use some combination of `sink()` and saving plot outputs
2. Sweave
3. R Markdown and knitr
Example Workflow

Raw Data → data_manage.R (all derivations, labeling, etc. contained in this file) → Analysis Data Sets

prog01_<desc>.R (individual analysis file #01) → ...

prog##_<desc>.R (individual analysis file ##)

master_include.R (links all programs for reporting)

Report.Rmd
Example Workflow

Raw Data → data_manage.R (all derivations, labeling, etc. contained in this file) → Analysis Data Sets

prog01_<desc>.R (individual analysis file #01)

... → master_include.R (links all programs for reporting)

prog##_<desc>.R (individual analysis file ##)

Report.Rmd
Example Workflow

Raw Data → data_manage.R
(all derivations, labeling, etc. contained in this file) → Analysis Data Sets

prog01_<desc>.R
(individual analysis file #01)

... →

prog##_<desc>.R
(individual analysis file ##)

master_include.R
(links all programs for reporting)

Report.Rmd
Object-Oriented Markdown

```
Analysis Prog01
```

```
Analysis Prog02

.
.
.
```

```
Analysis ProgXX
```

```
Result Objects

 Results01

 Results02

 ResultsXX
```

```
Load saved Results
```

```
Markdown Document

Print (Results01)
Print (Results02)
.
Print (ResultsXX)
```
T-test revisited

# Descriptive statistics
means_sds <- dataset %>% group_by(group) %>%
    summarise(Means=mean(outcome), sd=sd(outcome),
                se=sd/sqrt(n()), min=min(outcome), max=max(outcome))

# T-test results
ttest_out <- t.test(outcome ~ group, data=dataset)

# Histogram
plot1 <- ggplot(dataset, aes(outcome, fill=group)) +
    geom_histogram(color="black", bins=10) +
    facet_wrap(~group, ncol=1)

# Boxplot
plot2 <- ggplot(dataset, aes(Group, outcome)) +
    geom_boxplot() + coord_flip()

# Save results in output object
output_object <- list(means_sds, ttest_out, plot1, plot2)
T-test revisited

```r
print(output_object[[1]])
```

```
## # A tibble: 2 x 6
## #   group Means     sd     se  min  max
## 1 Group A -0.0378587 1.0410510 0.14722685 -2.3525414 2.541753
## 2 Group B  0.46369624 0.5264189 0.07444688 -0.5948096 1.612442
print(output_object[[2]])
```

```
## Welch Two Sample t-test
##
## data:  outcome by group
## t = -3.04, df = 72.52, p-value = 0.00329
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.8303715 -0.1726927
## sample estimates:
## mean in group Group A mean in group Group B
## -0.0378587 0.46369624
print(output_object[[3]])
```
Summary

- Approach decouples coding from printing in R Markdown
- Still allows reproducible research
- Allows for easy/quick report generation
- Supports team science
- Can still leverage additional markdown capabilities