

Booklet of Code and Output
for
STAC32 Midterm Exam

List of Figures in this document by page:

List of Figures

1	Packages	2
2	Texas highway mowing data	3
3	Uptake rate data	4
4	Histogram of uptake rates	5
5	T -test output	5
6	95% confidence interval for population mean uptake rate	6
7	Output from power estimation analysis	6
8	Power curve for different sample sizes	7
9	Histogram of wrist-extension data	8
10	T-test for wrist-extension data	9
11	Sign test for wrist-extension data	9
12	Example of “filter”	10
13	Further analysis for wrist extension data	11
14	Chess study data	12
15	HSAM data	12
16	HSAM overall median test score	13
17	HSAM Mood’s median test	13
18	HSAM facettted normal quantile plots	14
19	Penetrability data (randomly chosen rows)	15
20	Penetrability data boxplot	16
21	Penetrability data facettted normal quantile plots	17
22	Penetrability data ANOVA	18
23	Penetrability data Welch ANOVA	19
24	Penetrability data Mood’s median test	20

```
library(tidyverse)

## -- Attaching packages -----
tidyverse 1.2.1 --
## v ggplot2 3.2.1    v purrr  0.3.2
## v tibble  2.1.3    v dplyr  0.8.3
## v tidyr   1.0.0    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0
## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(smmr)
library(PMCMRplus)
```

Figure 1: Packages

height	frequency	vegetation
0.05m	1-per-year	17.3
0.05m	1-per-year	19.3
0.05m	1-per-year	15
0.05m	1-per-year	16.7
0.10m	1-per-year	16
0.10m	1-per-year	15.6
0.10m	1-per-year	16.9
0.10m	1-per-year	15
0.20m	1-per-year	16.7
0.20m	1-per-year	17.9
0.20m	1-per-year	15.9
0.20m	1-per-year	13.7
0.05m	2-per-year	22.4
0.05m	2-per-year	20.8
0.05m	2-per-year	24.5
0.05m	2-per-year	21.7
0.10m	2-per-year	23.9
0.10m	2-per-year	23.6
0.10m	2-per-year	21.7
0.10m	2-per-year	23.8
0.20m	2-per-year	24.7
0.20m	2-per-year	26.3
0.20m	2-per-year	27.2
0.20m	2-per-year	26.4
0.05m	3-per-year	18.6
0.05m	3-per-year	17.9
0.05m	3-per-year	16.1
0.05m	3-per-year	19.4
0.10m	3-per-year	22.2
0.10m	3-per-year	25.6
0.10m	3-per-year	21.8
0.10m	3-per-year	23.6
0.20m	3-per-year	27
0.20m	3-per-year	25.3
0.20m	3-per-year	23.8
0.20m	3-per-year	28

Figure 2: Texas highway mowing data

uptake
7251
6871
9632
6866
9094
5849
8957
7978
7064
7494
7883
8178
7523
8724
7468

Figure 3: Uptake rate data

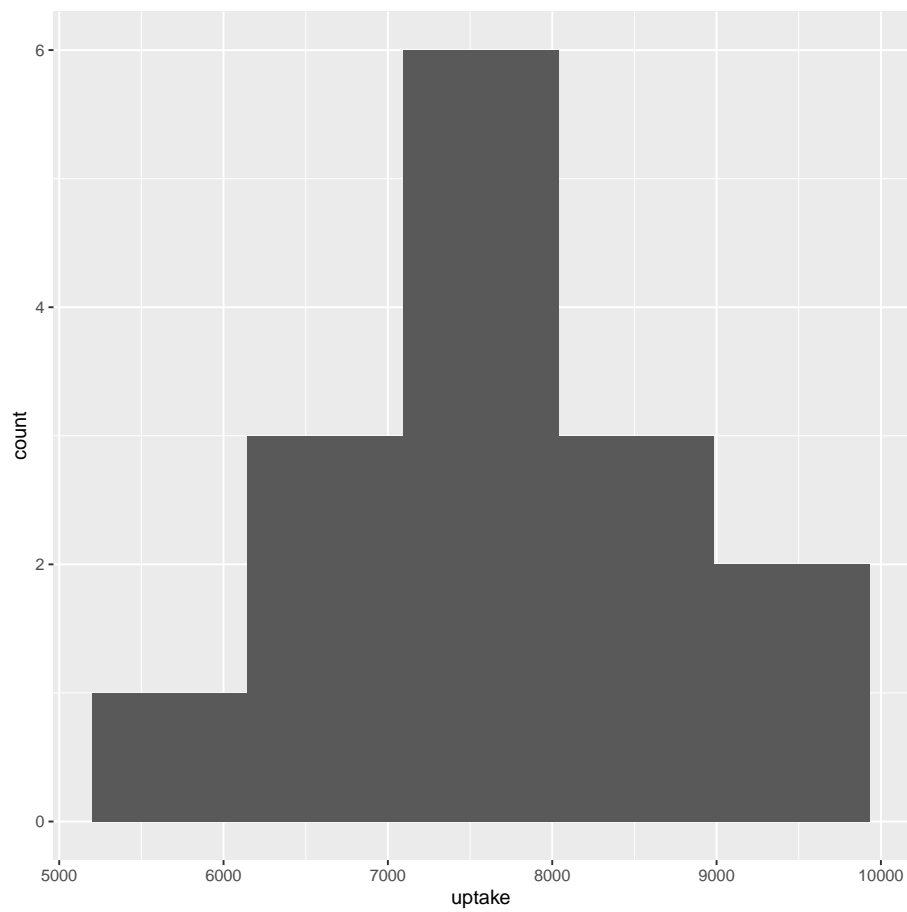


Figure 4: Histogram of uptake rates

```
##
## One Sample t-test
##
## data: uptake
## t = -0.81599, df = 14, p-value = 0.2141
## alternative hypothesis: true mean is less than 8000
## 95 percent confidence interval:
##      -Inf 8244.674
## sample estimates:
## mean of x
##      7788.8
```

Figure 5: T -test output

```
##
## One Sample t-test
##
## data: uptake
## t = 30.093, df = 14, p-value = 3.998e-14
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 7233.672 8343.928
## sample estimates:
## mean of x
## 7788.8
```

Figure 6: 95% confidence interval for population mean uptake rate

```
## # A tibble: 2 x 2
##   `pvals < 0.05`      n
##   <lgl>             <int>
## 1 FALSE             587
## 2 TRUE              413
```

Figure 7: Output from power estimation analysis

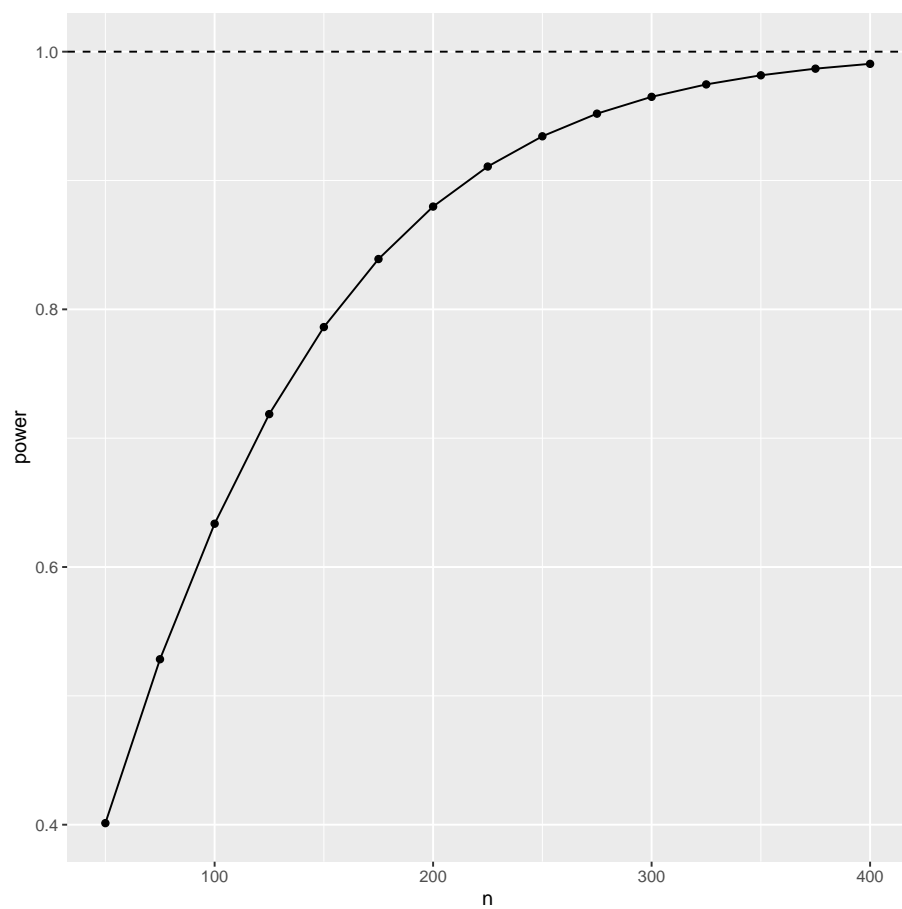


Figure 8: Power curve for different sample sizes


```
ggplot(wrist, aes(x=extension)) + geom_histogram(bins=10)
```

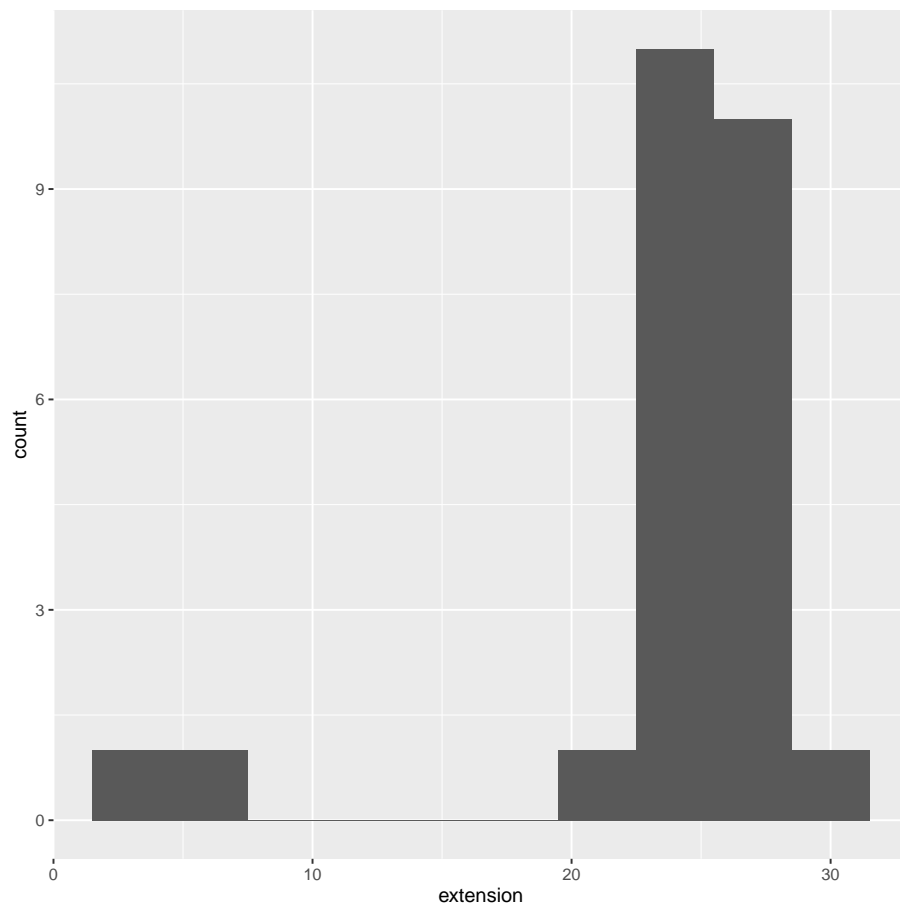


Figure 9: Histogram of wrist-extension data

```

t.test(wrist$extension, mu=24, alternative="greater")

##
## One Sample t-test
##
## data: wrist$extension
## t = 0.2011, df = 24, p-value = 0.4212
## alternative hypothesis: true mean is greater than 24
## 95 percent confidence interval:
## 22.19819      Inf
## sample estimates:
## mean of x
## 24.24

```

Figure 10: T-test for wrist-extension data

```

sign_test(wrist, extension, 24)

## $above_below
## below above
##      3      17
##
## $p_values
## alternative      p_value
## 1          lower 0.999798775
## 2           upper 0.001288414
## 3      two-sided 0.002576828

```

Figure 11: Sign test for wrist-extension data

Suppose we have a data frame like this:

```
d
## # A tibble: 5 x 2
##       x g
##   <dbl> <chr>
## 1    10 a
## 2    11 b
## 3    12 a
## 4    13 b
## 5    14 a
```

Then we can select all the rows for which the column `g` is equal to the text `a` like this:

```
d %>% filter(g=="a")
## # A tibble: 3 x 2
##       x g
##   <dbl> <chr>
## 1    10 a
## 2    12 a
## 3    14 a
```

Figure 12: Example of “filter”

```

wrist %>% filter(extension>10) -> wrist2
t.test(wrist2$extension, mu=24, alternative="greater")

##
## One Sample t-test
##
## data: wrist2$extension
## t = 4.5189, df = 22, p-value = 8.492e-05
## alternative hypothesis: true mean is greater than 24
## 95 percent confidence interval:
## 25.15915      Inf
## sample estimates:
## mean of x
## 25.86957

sign_test(wrist2, extension, 24)

## $above_below
## below above
##      1      17
##
## $p_values
##      alternative      p_value
## 1      lower 9.999962e-01
## 2      upper 7.247925e-05
## 3 two-sided 1.449585e-04

```

Figure 13: Further analysis for wrist extension data

```
## # A tibble: 12 x 4
##   student pre_test post_test difference
##   <dbl>   <dbl>   <dbl>   <dbl>
## 1     1     510     850     340
## 2     2     610     790     180
## 3     3     640     850     210
## 4     4     675     775     100
## 5     5     600     700     100
## 6     6     550     775     225
## 7     7     610     700     90
## 8     8     625     850     225
## 9     9     450     690     240
## 10    10     720     775     55
## 11    11     575     540    -35
## 12    12     675     680     5
```

Figure 14: Chess study data

```
hsam

## # A tibble: 29 x 2
##   memory test_score
##   <chr>      <dbl>
## 1 control     4
## 2 control     4
## 3 control     3
## 4 hsam        6
## 5 control     4
## 6 control     6
## 7 control     5
## 8 hsam        7
## 9 hsam        6
## 10 hsam       4
## # ... with 19 more rows
```

Figure 15: HSAM data

```
hsam %>% summarize(m=median(test_score))

## # A tibble: 1 x 1
##       m
##   <dbl>
## 1     4
```

Figure 16: HSAM overall median test score

```
## $table
##           above
## group    above below
##   control     5    10
##   hsam        8     0
##
## $test
##           what      value
## 1 statistic 9.435897436
## 2          df 1.000000000
## 3    P-value 0.002127789
```

Figure 17: HSAM Mood's median test

```
ggplot(hsam, aes(sample=test_score)) +  
  stat_qq() + stat_qq_line() + facet_wrap(~memory)
```

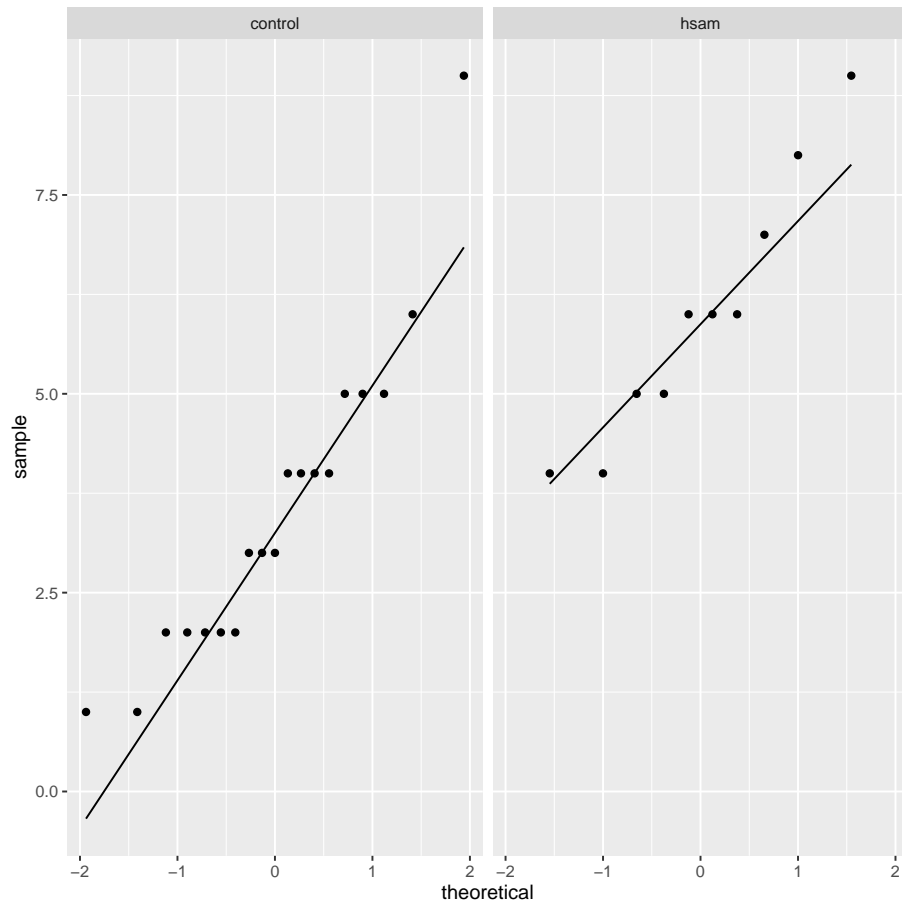


Figure 18: HSAM facetted normal quantile plots

```
## Parsed with column specification:
## cols(
##   plot = col_character(),
##   penetrability = col_double()
## )

## # A tibble: 20 x 2
##   plot penetrability
##   <chr>         <dbl>
## 1 B           3.13
## 2 C           4.91
## 3 B           3.26
## 4 A            3
## 5 B           3.86
## 6 A           2.9
## 7 A           2.86
## 8 C           3.99
## 9 B           3.38
## 10 A          3.18
## 11 C           4.3
## 12 C           3.94
## 13 A           2.92
## 14 A           2.86
## 15 B           3.38
## 16 C           4.2
## 17 A           2.96
## 18 B           3.02
## 19 C           4.34
## 20 A           2.78
```

Figure 19: Penetrability data (randomly chosen rows)

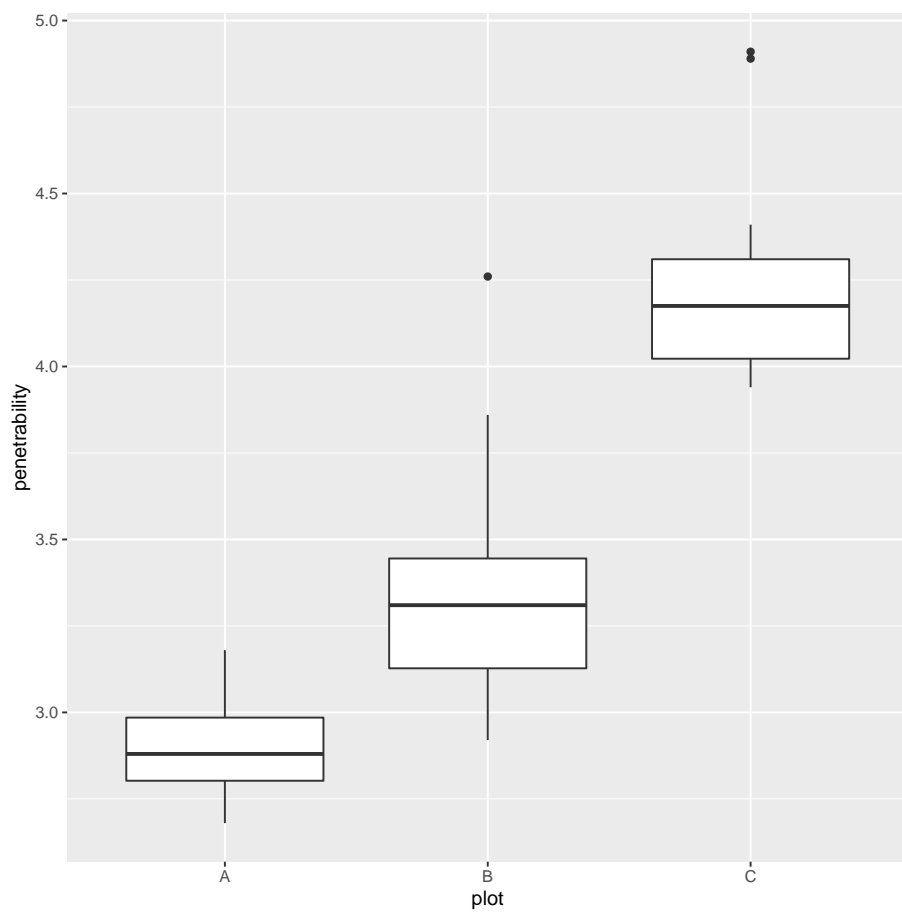
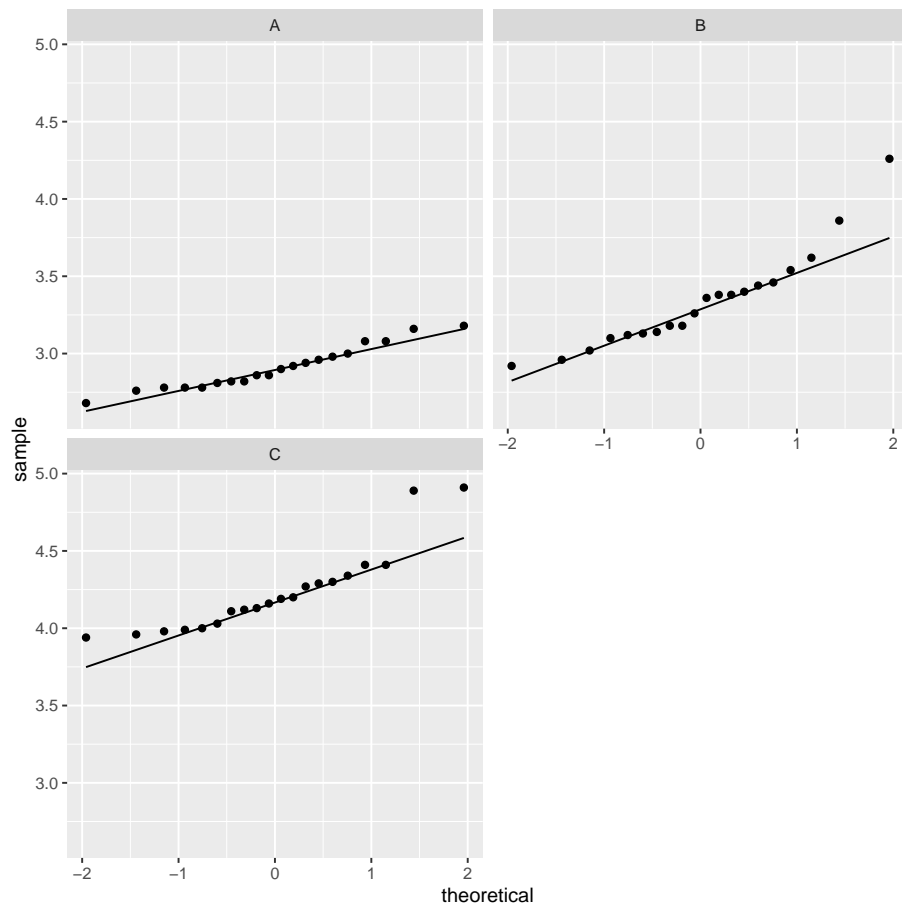


Figure 20: Penetrability data boxplot

```
ggplot(soil, aes(sample=penetrability)) + stat_qq() +
  stat_qq_line() + facet_wrap(~plot, ncol=2)
```



The purpose of the `ncol=2` is to arrange the plots as three cells of a 2×2 grid. By default, the three plots will come out in one row, side by side, which is harder to read.

Figure 21: Penetrability data faceted normal quantile plots

Part (i):

```
soil.1=aov(penetrability~plot, data=soil)
summary(soil.1)

##              Df Sum Sq Mean Sq F value Pr(>F)
## plot           2 18.260    9.130   140.5 <2e-16 ***
## Residuals     57  3.703    0.065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part (ii):

```
TukeyHSD(soil.1)

##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = penetrability ~ plot, data = soil)
##
## $plot
##      diff      lwr      upr    p adj
## B-A 0.428 0.2340375 0.6219625 5.6e-06
## C-A 1.324 1.1300375 1.5179625 0.0e+00
## C-B 0.896 0.7020375 1.0899625 0.0e+00
```

Figure 22: Penetrability data ANOVA

Part (i):

```
oneway.test(penetrability~plot, data=soil)

##
## One-way analysis of means (not assuming equal variances)
##
## data: penetrability and plot
## F = 186.86, num df = 2.000, denom df = 33.243, p-value < 2.2e-
16
```

Part (ii):

```
gamesHowellTest(penetrability~factor(plot), data=soil)

##
## Pairwise comparisons using Games-Howell test
## data: penetrability by factor(plot)
##
## A B
## B 2.7e-05 -
## C < 2e-16 4.5e-11
##
## P value adjustment method: none
## alternative hypothesis: two.sided
```

Figure 23: Penetrability data Welch ANOVA

Part (i):

```
median_test(soil, penetrability, plot)

## $table
##      above
## group above below
##    A      0    20
##    B     10    10
##    C     20     0
##
## $test
##      what      value
## 1 statistic 4.000000e+01
## 2          df 2.000000e+00
## 3    P-value 2.061154e-09
```

Part (ii):

```
pairwise_median_test(soil, penetrability, plot)

## # A tibble: 3 x 4
##   g1    g2    p_value adj_p_value
##   <chr> <chr>    <dbl>    <dbl>
## 1 A     B     5.40e- 6    1.62e- 5
## 2 A     C     2.54e-10    7.62e-10
## 3 B     C     1.25e- 8    3.76e- 8
```

Figure 24: Penetrability data Mood's median test