Booklet of Figures
for
STAC33 Midterm Exam

You may take away this booklet after the exam, and therefore you are free to tear off pages as you wish.

List of Figures in this document by page:

List of Figures

```
library(tidyverse)

## -- Attaching packages ---------------------------------------
tidyverse 1.3.0 --
## v ggplot2 3.2.1     v purrr   0.3.3
## 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)
```

Figure 1: Packages

community	station	price
Langley	Chevron	93.9
Langley	Chevron	101.9
Langley	Chevron	101.9
Langley	Esso	101.9
Langley	Esso	93.9
Langley	Esso	101.9
Langley	Shell	104.2
Langley	Shell	104.9
Langley	Shell	101.9
Surrey	Chevron	101.9
Surrey	Chevron	104.9
Surrey	Chevron	101.9
Surrey	Esso	103.2
Surrey	Esso	101.9
Surrey	Esso	105.2
Surrey	Shell	101.9
Surrey	Shell	105.2
Surrey	Shell	105.5
Abbotsford	Chevron	92.5
Abbotsford	Chevron	92.5
Abbotsford	Chevron	98.5
Abbotsford	Esso	89.0
Abbotsford	Esso	88.9
Abbotsford	Esso	92.5
Abbotsford	Shell	92.5
Abbotsford	Shell	88.9
Abbotsford	Shell	87.9

Figure 2: Vancouver-area gas data

```
## Parsed with column specification:
## cols(
##  community = col_character(),
##  station = col_character(),
##  price = col_double()
## )
```

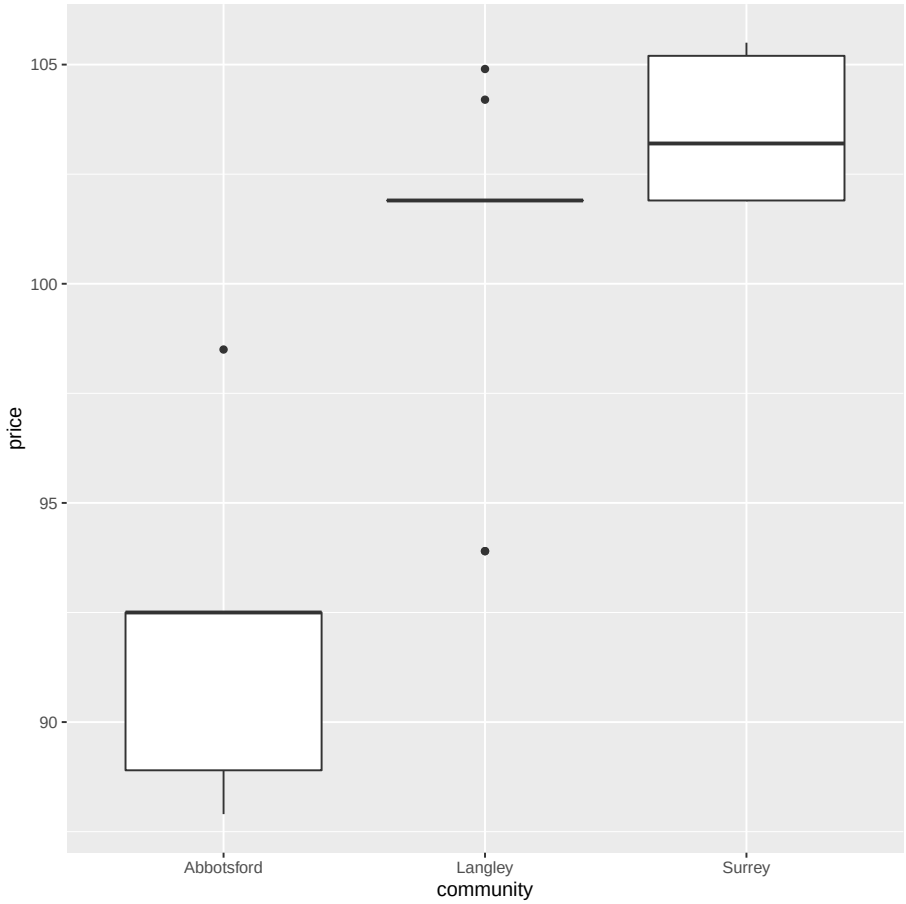

Figure 3: Gas data boxplot

```
weights
```

```
## # A tibble: 16 x 1
##    weight
##     <int>
##  1    173
##  2    178
##  3    145
##  4    146
##  5    157
##  6    175
##  7    173
##  8    137
##  9    152
## 10    171
## 11    163
## 12    170
## 13    135
## 14    165
## 15    199
## 16    131
```

Figure 4: Adult male weight data

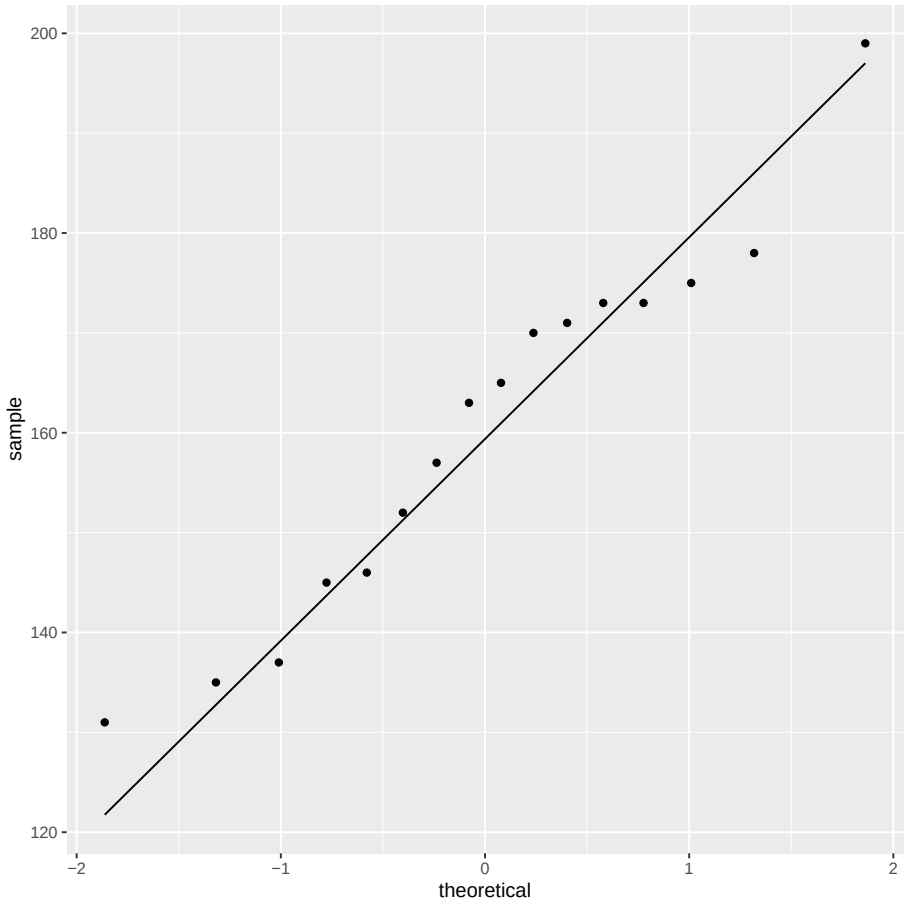

Figure 5: Weights normal quantile plot

```
##
##  One Sample t-test
##
## data:  weight
## t = 0.13498, df = 15, p-value = 0.4472
## alternative hypothesis: true mean is greater than 160
## 95 percent confidence interval:
##  152.5077      Inf
## sample estimates:
## mean of x
##   160.625
```

Figure 6: Weights t-test output

```
summary(weights)
```

```
##      weight
##  Min.   :131.0
##  1st Qu.:145.8
##  Median :164.0
##  Mean   :160.6
##  3rd Qu.:173.0
##  Max.   :199.0
```

Figure 7: Weights data frame summary

```
## Parsed with column specification:
## cols(
##  method = col_character(),
##  heat_change = col_double()
## )
```

```
##        method heat_change
## 1     mixture       79.98
## 2     mixture       80.04
## 3     mixture       80.02
## 4     mixture       80.04
## 5     mixture       80.03
## 6     mixture       80.03
## 7     mixture       80.04
## 8     mixture       79.97
## 9     mixture       80.05
## 10    mixture       80.03
## 11    mixture       80.02
## 12    mixture       80.02
## 13    mixture       80.02
## 14 electrical       80.02
## 15 electrical       79.94
## 16 electrical       79.98
## 17 electrical       79.97
## 18 electrical       79.97
## 19 electrical       80.03
## 20 electrical       79.95
## 21 electrical       79.97
```

Figure 8: Fusion data

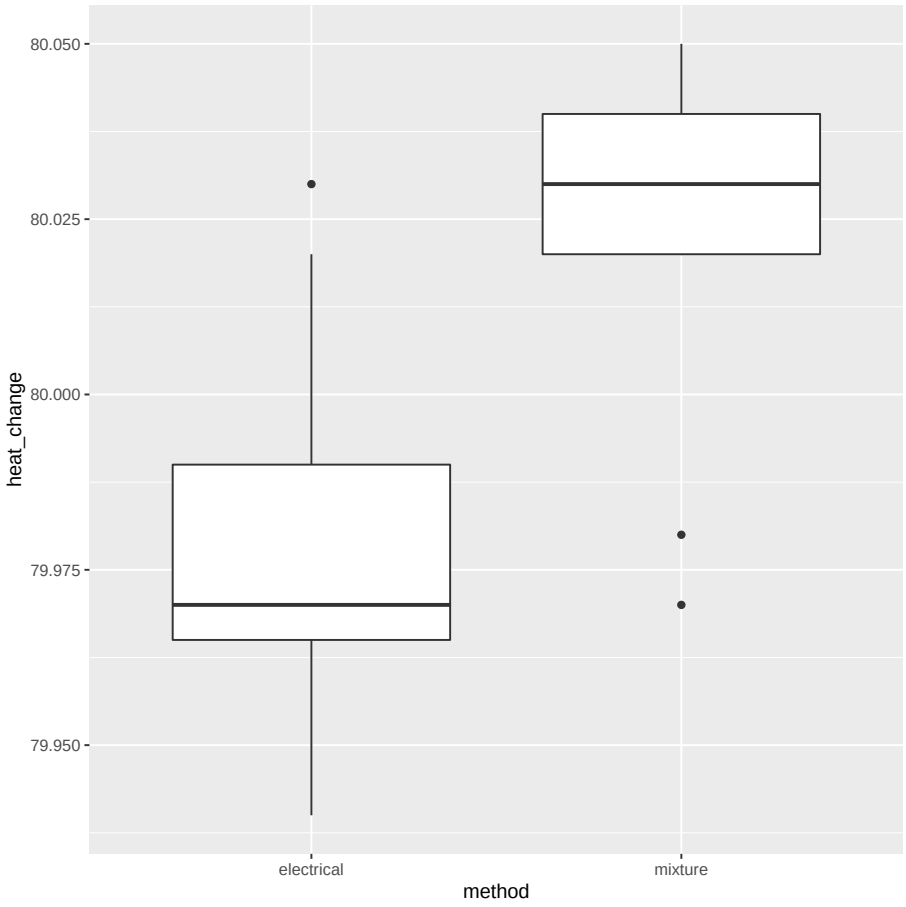

Figure 9: Fusion boxplot

```r
t.test(heat_change~method, data=fusion)
```

```
##
##  Welch Two Sample t-test
##
## data:  heat_change by method
## t = -3.3991, df = 11.711, p-value = 0.005454
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.07155483 -0.01556055
## sample estimates:
## mean in group electrical    mean in group mixture
##                 79.97875                 80.02231
```

Figure 10: Fusion analysis 1

```r
t.test(heat_change~method, data=fusion, var.equal=T)
```

```
##
##  Two Sample t-test
##
## data:  heat_change by method
## t = -3.6612, df = 19, p-value = 0.00166
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.06845855 -0.01865684
## sample estimates:
## mean in group electrical    mean in group mixture
##                 79.97875                 80.02231
```

Figure 11: Fusion analysis 2

```
median_test(fusion, heat_change, method)
```

```
## $table
##             above
## group        above below
##   electrical     1     6
##   mixture        7     2
##
## $test
##        what      value
## 1 statistic 6.34920635
## 2        df 1.00000000
## 3   P-value 0.01174338
```

Figure 12: Fusion analysis 3

```r
tibble(x=seq(0, 10, 0.1)) %>%
  mutate(density=dexp(x, 0.5)) %>%
  ggplot(aes(x=x, y=density)) + geom_line()
```

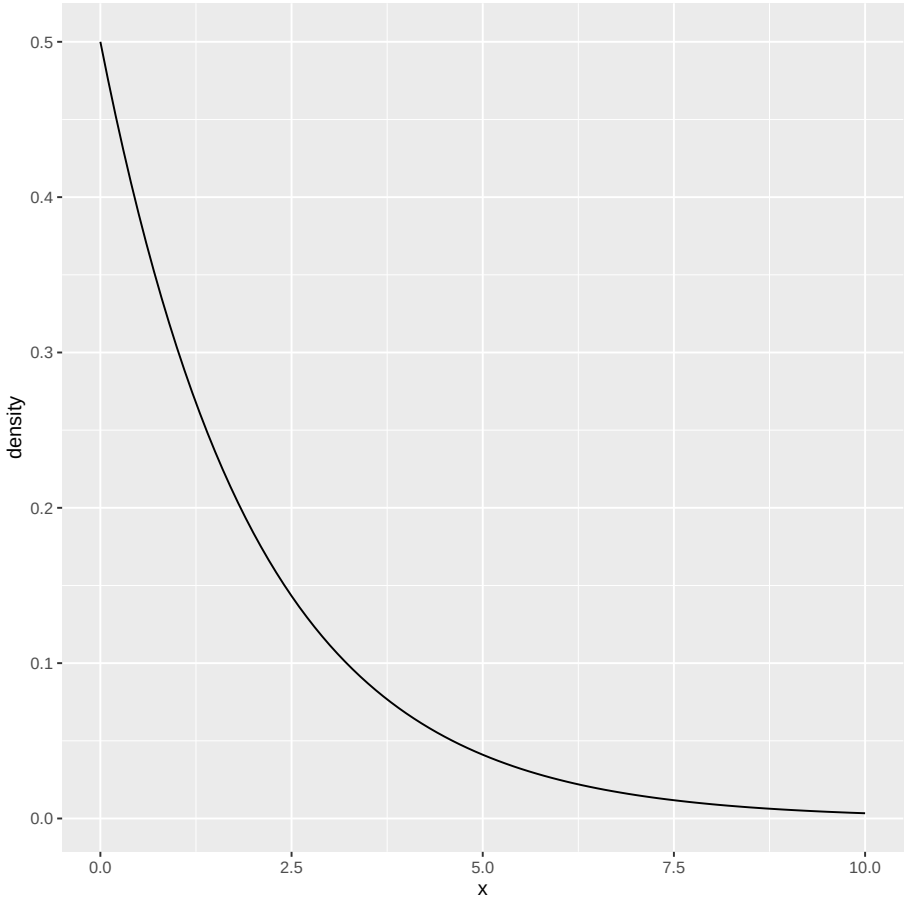

Figure 13: Exponential distribution density function, $\lambda = 0.5$

```r
v <- c(3,4,5,6,7)
vv <- tibble(v)
vv
```

```
## # A tibble: 5 x 1
##       v
##   <dbl>
## 1     3
## 2     4
## 3     5
## 4     6
## 5     7
```

```r
sign_test(vv, v, 2.5)
```

```
## $above_below
## below above
##     0     5
##
## $p_values
##   alternative p_value
## 1       lower 1.00000
## 2       upper 0.03125
## 3   two-sided 0.06250
```

```r
pval_sign0(2.5, v)
```

```
## [1] 0.0625
```

Figure 14: Example of pval-sign0

```
## # A tibble: 24 x 2
##    city_size  arrest_rate
##    <chr>            <dbl>
##  1 large_city          45
##  2 small_city          23
##  3 suburb              25
##  4 rural                8
##  5 large_city          34
##  6 small_city          18
##  7 suburb              17
##  8 rural               16
##  9 large_city          41
## 10 small_city          27
## 11 suburb              19
## 12 rural               14
## 13 large_city          42
## 14 small_city          21
## 15 suburb              28
## 16 rural               17
## 17 large_city          37
## 18 small_city          26
## 19 suburb              31
## 20 rural               10
## 21 large_city          28
## 22 small_city          34
## 23 suburb              37
## 24 rural               23
```

Figure 15: Narcotics arrest rate data

```
ggplot(narc, aes(x=city_size, y=arrest_rate)) + geom_boxplot()
```

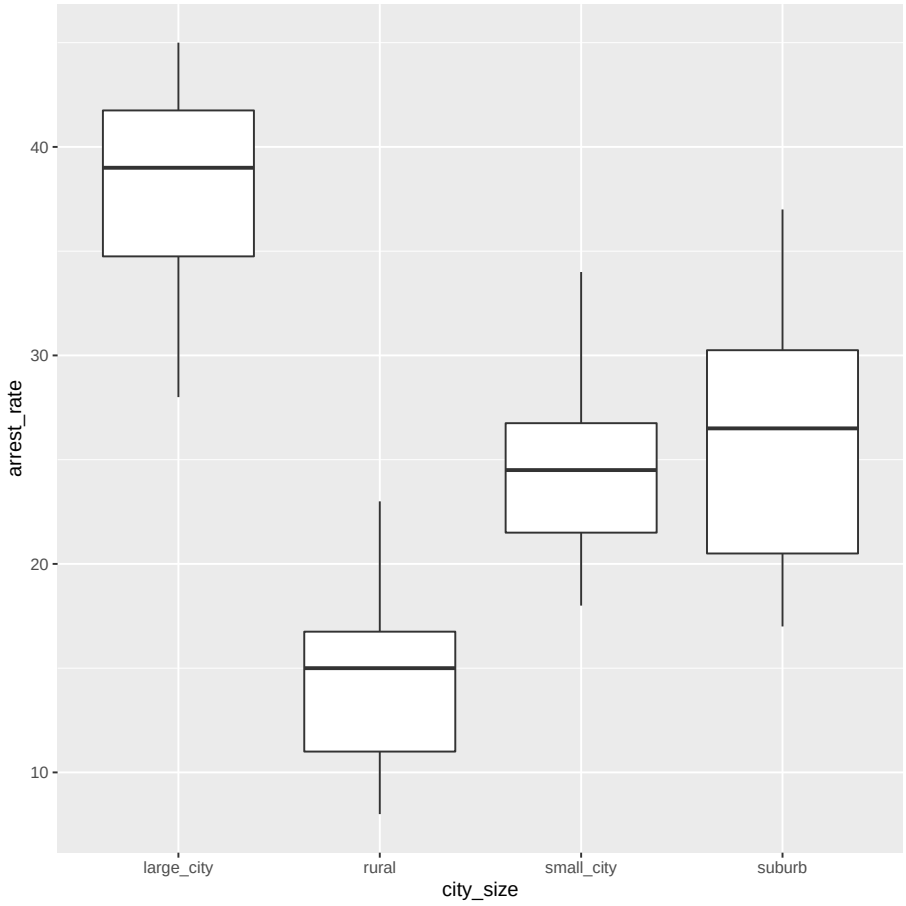

Figure 16: Narcotics arrest rate boxplot

```
narc.1=aov(arrest_rate~city_size, data=narc)
summary(narc.1)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## city_size    3 1618.8   539.6   14.02 3.76e-05 ***
## Residuals   20  769.8    38.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 17: Narcotics arrest rate ANOVA

```
TukeyHSD(narc.1)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = arrest_rate ~ city_size, data = narc)
##
## $city_size
##                              diff         lwr        upr     p adj
## rural-large_city      -23.166667 -33.1923901 -13.140943 0.0000147
## small_city-large_city -13.000000 -23.0257235  -2.974277 0.0083460
## suburb-large_city     -11.666667 -21.6923901  -1.640943 0.0189865
## small_city-rural       10.166667   0.1409432  20.192390 0.0461249
## suburb-rural           11.500000   1.4742765  21.525723 0.0210020
## suburb-small_city       1.333333  -8.6923901  11.359057 0.9818944
```

Figure 18: Narcotics arrest rate Tukey

```
narc.2=oneway.test(arrest_rate~city_size, data=narc)
narc.2

##
##  One-way analysis of means (not assuming equal variances)
##
## data:  arrest_rate and city_size
## F = 14.391, num df = 3.000, denom df = 11.042, p-value = 0.0003932

library(PMCMRplus)
gamesHowellTest(arrest_rate~factor(city_size), data=narc)

##
## Pairwise comparisons using Games-Howell test
## data:  arrest_rate by factor(city_size)

##            large_city rural   small_city
## rural      0.00021    -       -
## small_city 0.01499    0.03848 -
## suburb     0.06194    0.05474 0.98437

##
## P value adjustment method:  none
## alternative hypothesis:  two.sided
```

Figure 19: Narcotics arrest rate analysis 2