Univariate - MA Data Analysis

Akihiko Mori

Univariable —

Open days

```
##   obs_days open_days closed_days
## 1      169         8         161
```

```
## # A tibble: 2 x 3
##   is_closed     n  prop
##   <lgl>     <int> <dbl>
## 1 FALSE       161  95.3
## 2 TRUE          8   4.7
```

Basic Summary of Dependent Variables

```
## # A tibble: 4 x 13
##   variable        n   min   max median    q1    q3   iqr   mad  mean    sd    se
##   <fct>       <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 food_loss_~   161     0 13.8    7.35  6.7   8.4   1.7  1.11  7.83  2.17  0.171
## 2 food_waste~   161     0  6.55   2.1   1.1   2.95  1.85 1.33  2.19  1.40  0.111
## 3 liquid_was~   161     0  4.5    1.5   0.65  2.05  1.4  1.04  1.48  0.995 0.078
## 4 solid_wast~   161     0  2.95   0.65  0.35  0.95  0.6  0.445 0.708 0.499 0.039
## # i 1 more variable: ci <dbl>
```

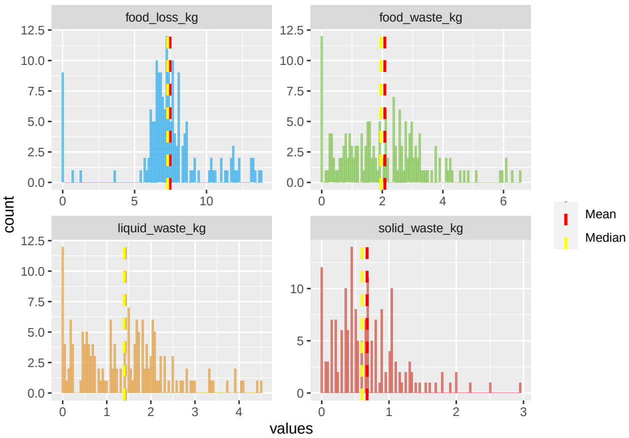
```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:rstatix':
##
##     select
```

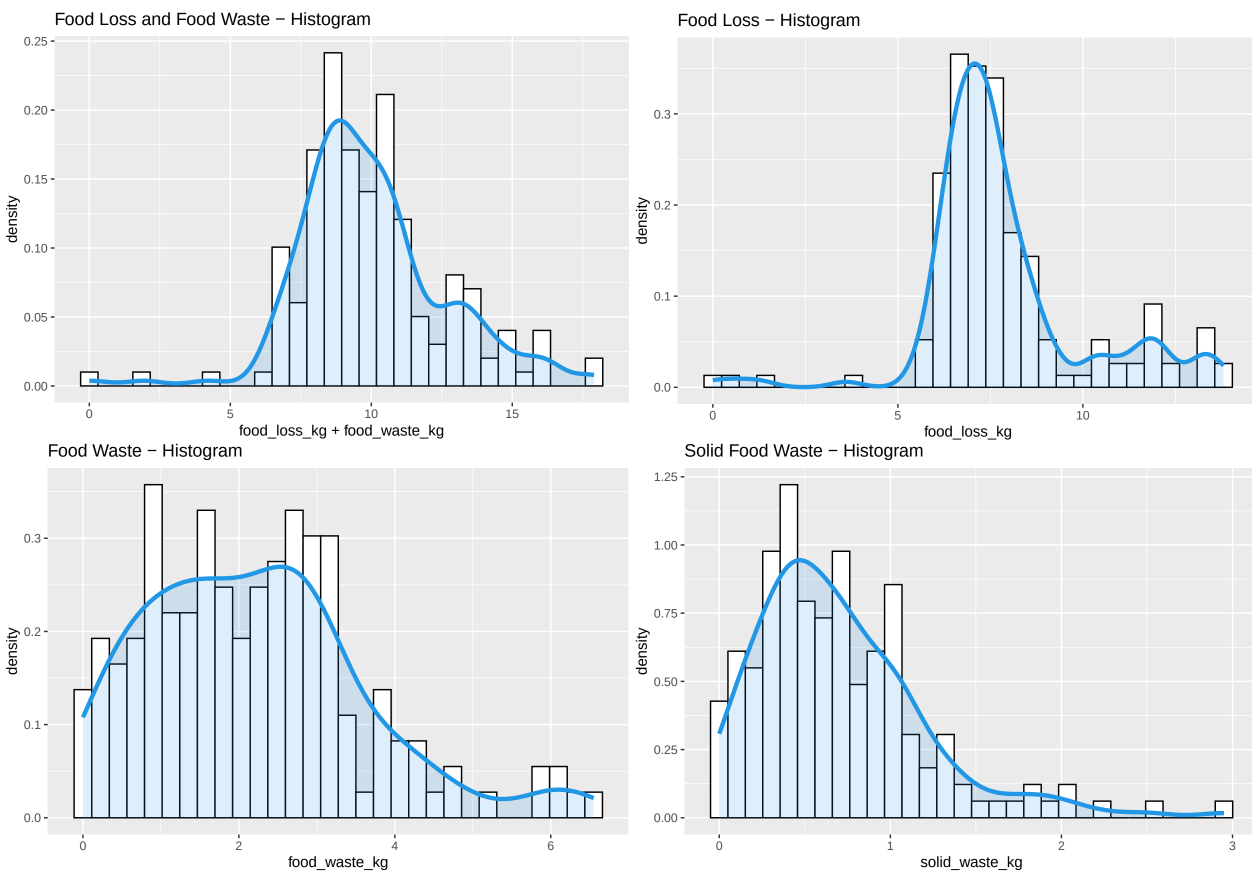
```
## The following object is masked from 'package:dplyr':
##
##     select
```

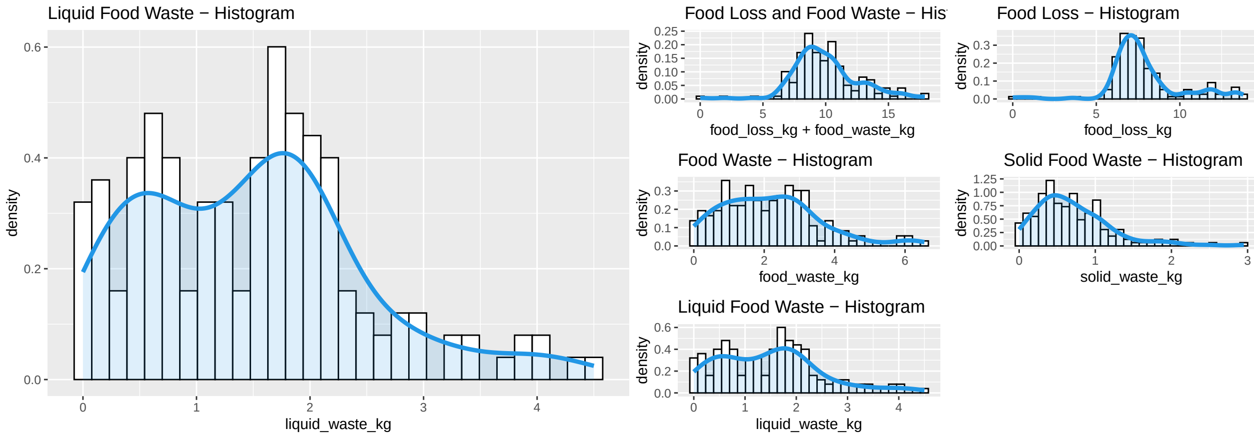
Histograms —

Normal histogram

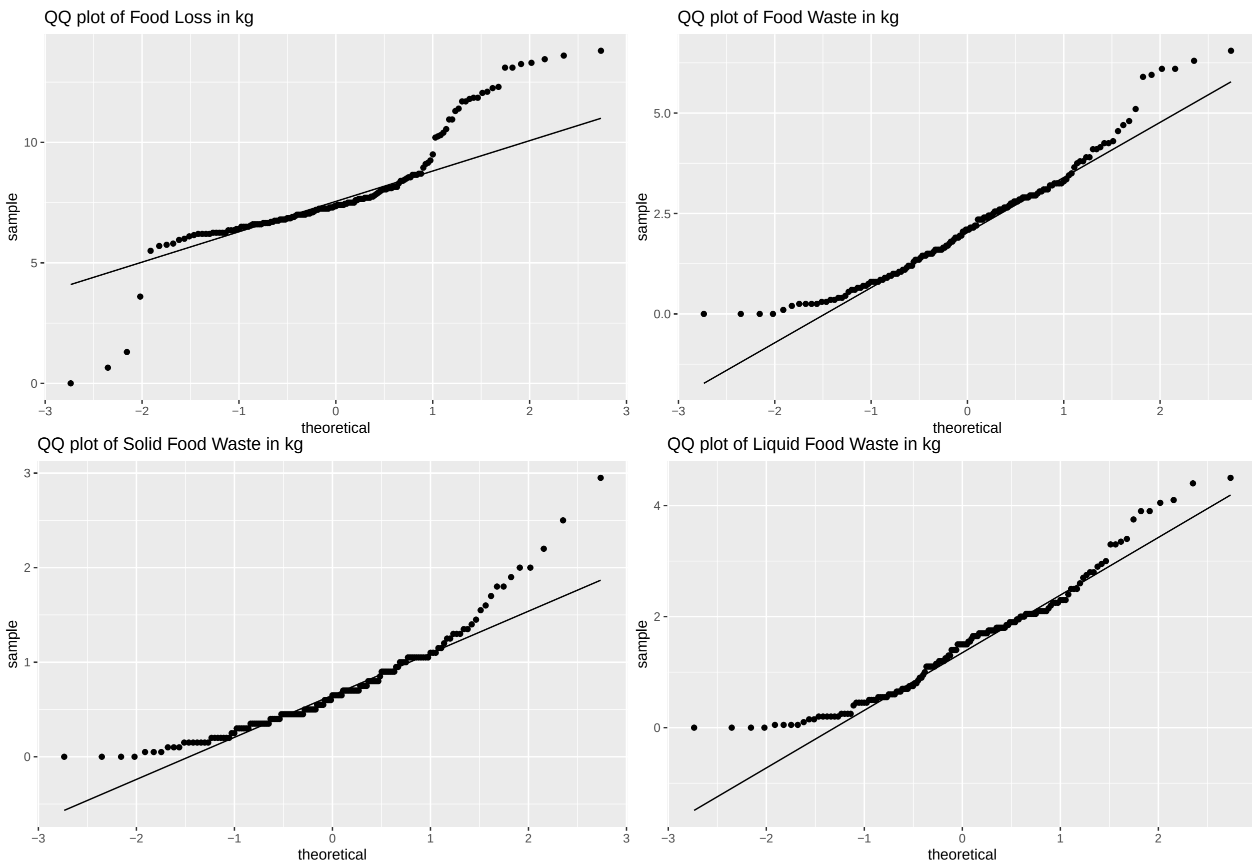


Histogram with density





Q-Q plot

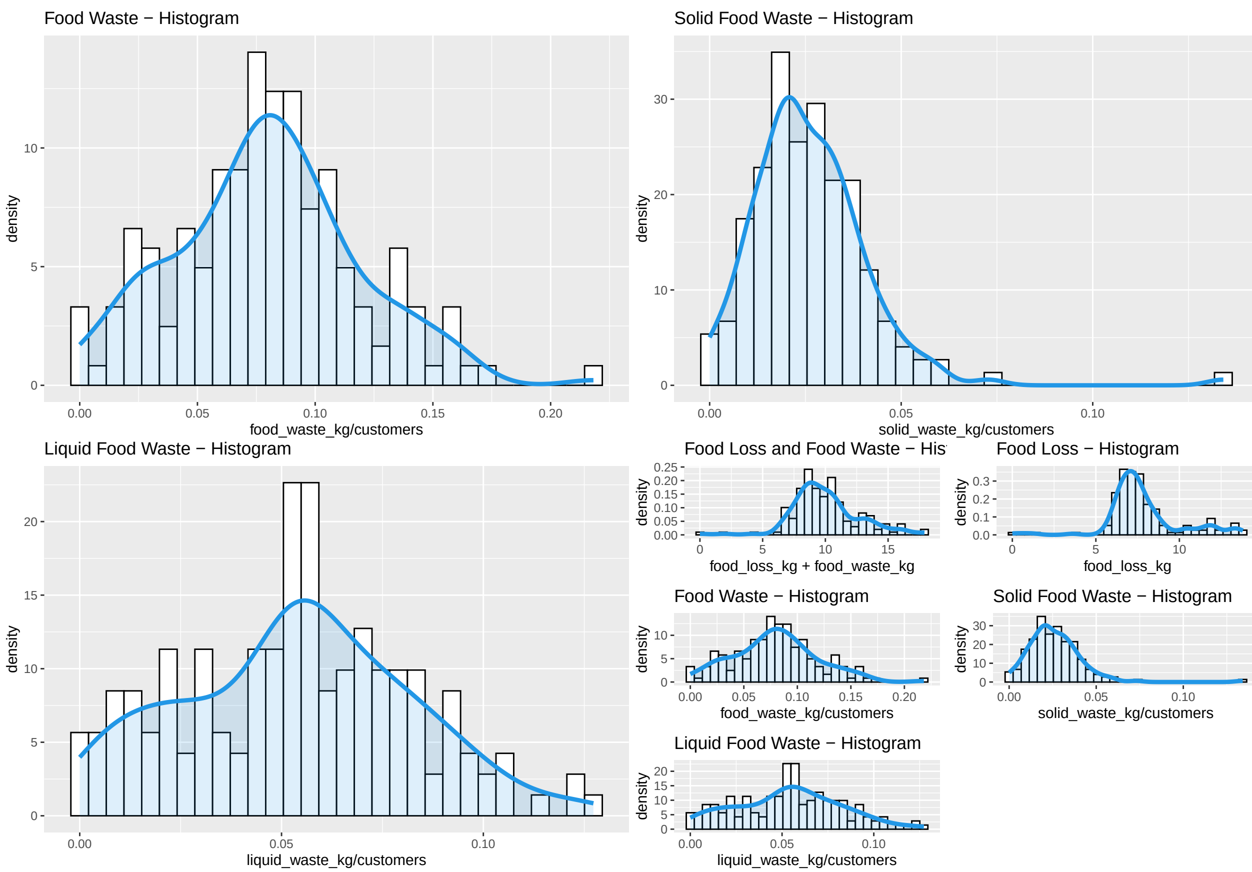


shapiro test

```
## # A tibble: 3 x 3
##   variable        statistic           p
##   <chr>               <dbl>       <dbl>
## 1 food_waste_kg       0.952 0.0000260
## 2 liquid_waste_kg     0.951 0.0000192
## 3 solid_waste_kg      0.903 0.00000000783
```

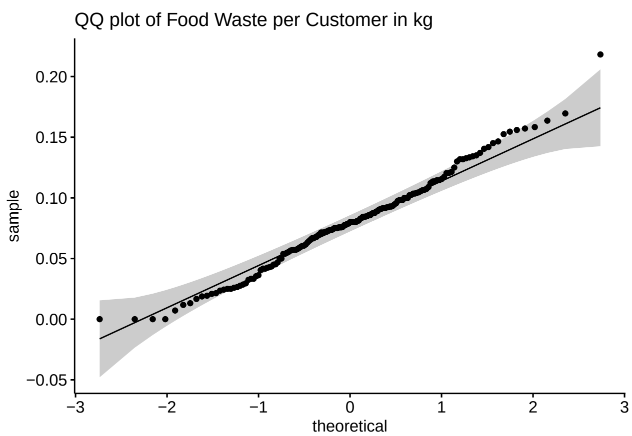
From the output, all the p-value is far less than 0.05; so implying that the distribution of the data are significantly different from normal distribution. In other words, we can not assume the normality.

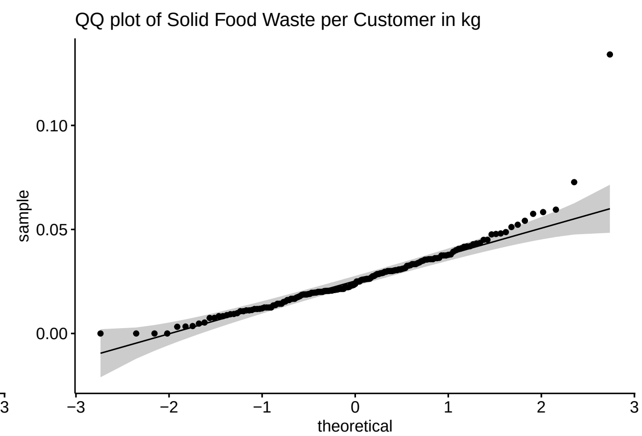
Histogram per customer

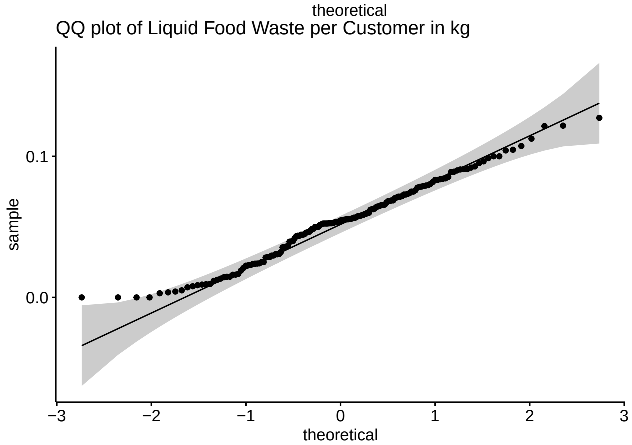


Q-Q plot per customer

```
##
## Attaching package: 'ggpubr'

## The following object is masked from 'package:forecast':
##
##     gghistogram
```







shapiro test for per customer

```
## # A tibble: 3 x 3
##   variable          statistic        p
##   <chr>                 <dbl>    <dbl>
## 1 food_waste_p_kg       0.987 1.38e- 1
## 2 liquid_waste_p_kg     0.984 6.10e- 2
## 3 solid_waste_p_kg      0.863 6.24e-11
```

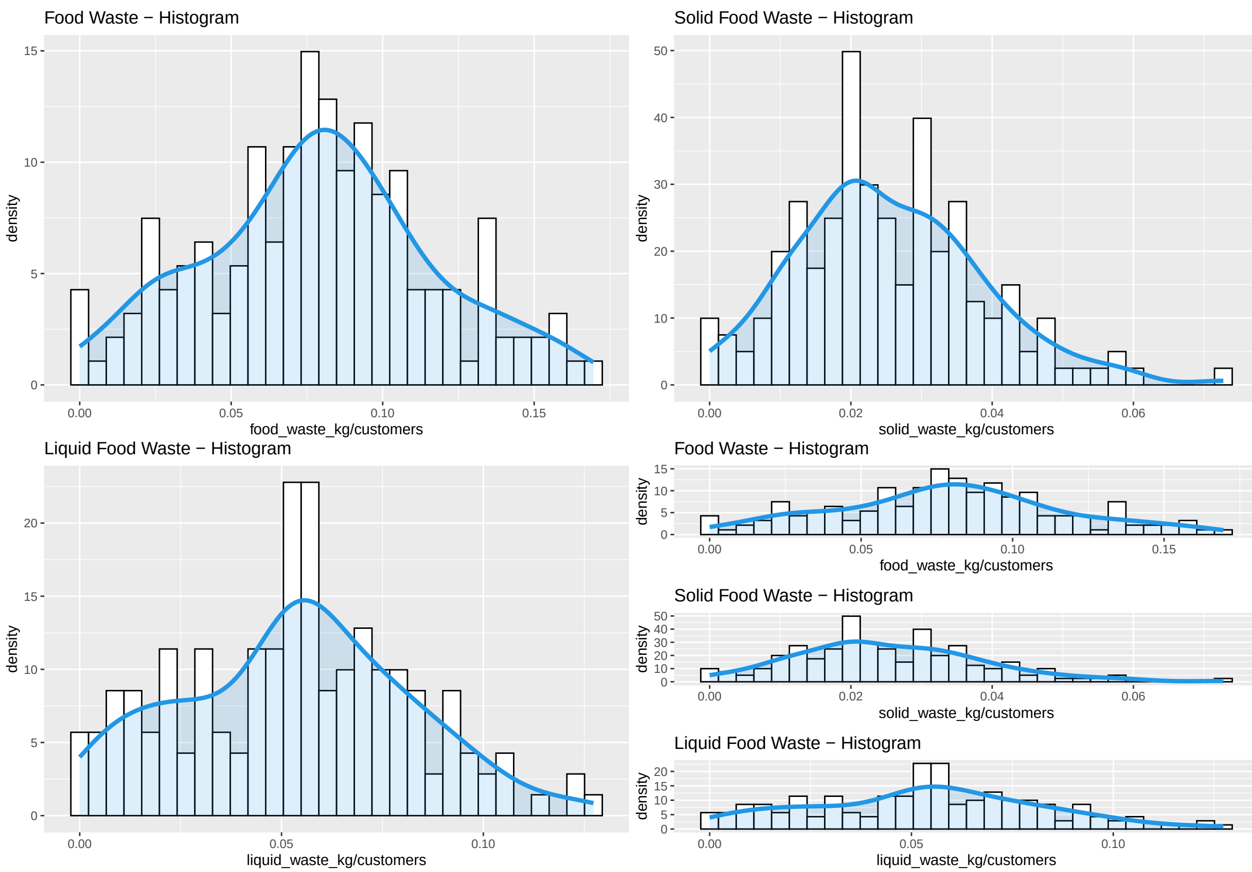
From the output, the p-value of solid food waste per customer is far less that the significant level of 0.05; but the others are not. So it imply that the distribution of the data for solid food waste per customer is significantly different from normal distribution. In other words, we can assume the normality for food waste and liquid food waste per customer but not for solid food waste.

Histogram per customer w/o outlier

```
## [1] 46
```

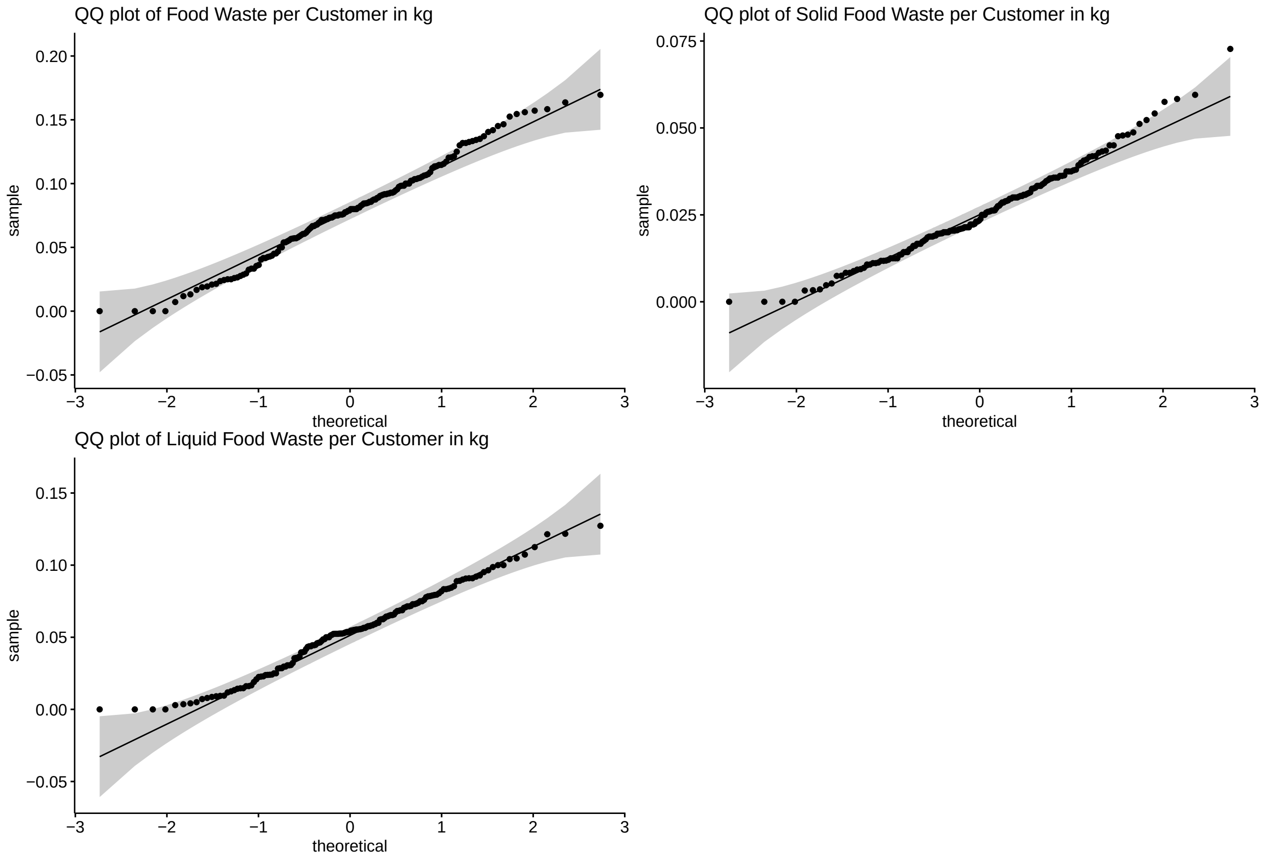
```
## [1] 46
```

```
## [1] "2022-11-08"
```



Q-Q plot per capita w/o outlier

```
##
## Attaching package: 'qqplotr'

## The following objects are masked from 'package:ggplot2':
##
##     stat_qq_line, StatQqLine
```

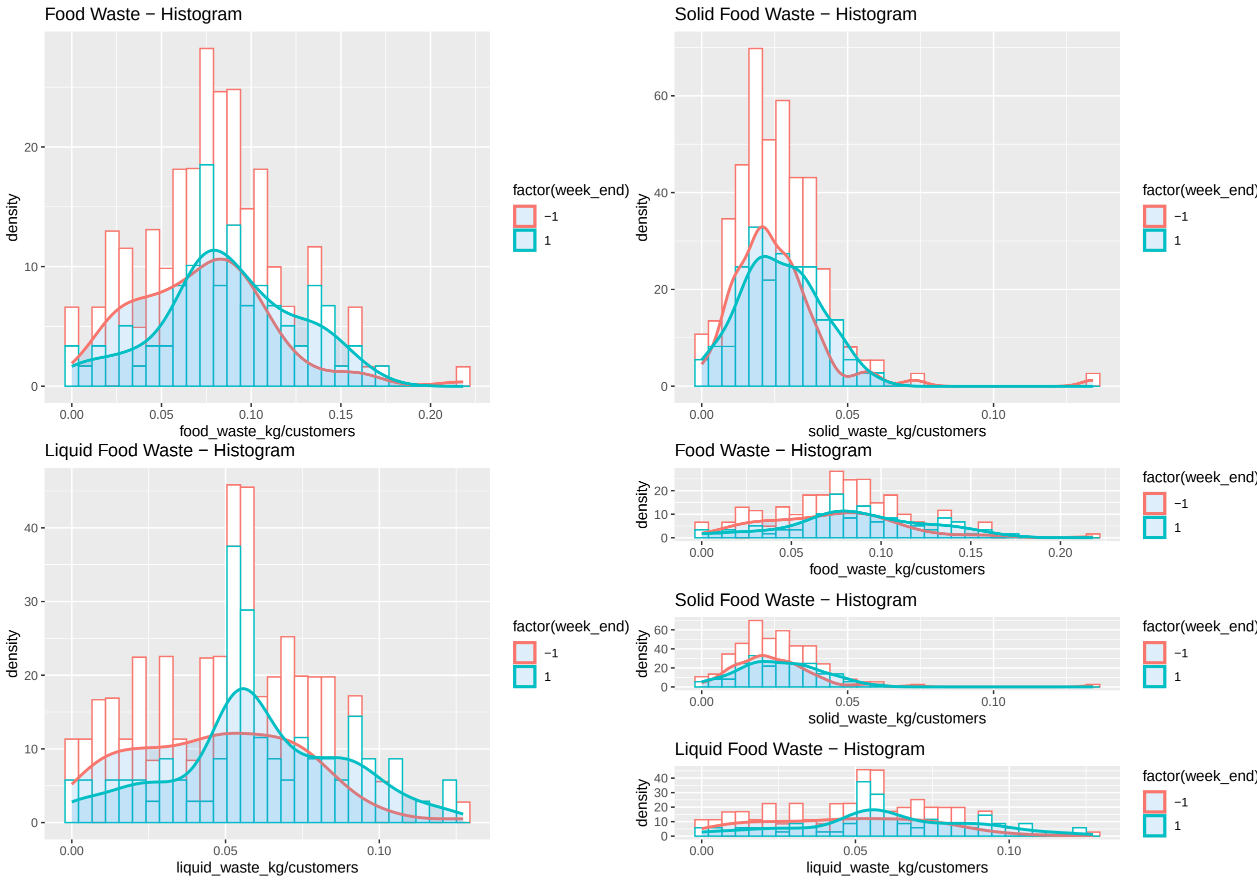


shapiro test for per capita w/o outlier

```
## # A tibble: 3 x 3
##   variable           statistic      p
##   <chr>                  <dbl>  <dbl>
## 1 food_waste_p_kg        0.988 0.210
## 2 liquid_waste_p_kg      0.984 0.0601
## 3 solid_waste_p_kg       0.980 0.0222
```

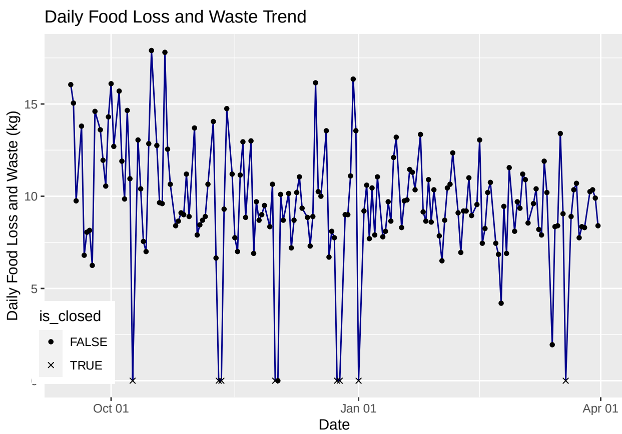
From the output, the p-value of solid food waste per customer is far less that the significant level of 0.05; but the others are not. So it imply that the distribution of the data for solid food waste per customer is significantly different from normal distribution. In other words, we can assume the normality for food waste and liquid food waste per customer but not for solid food waste.

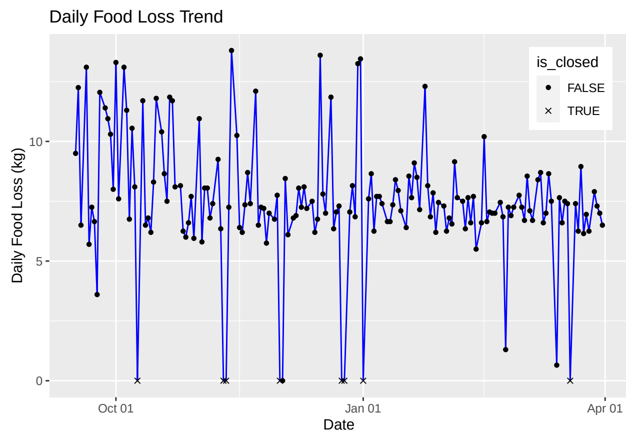
Histogram weekdays_ends

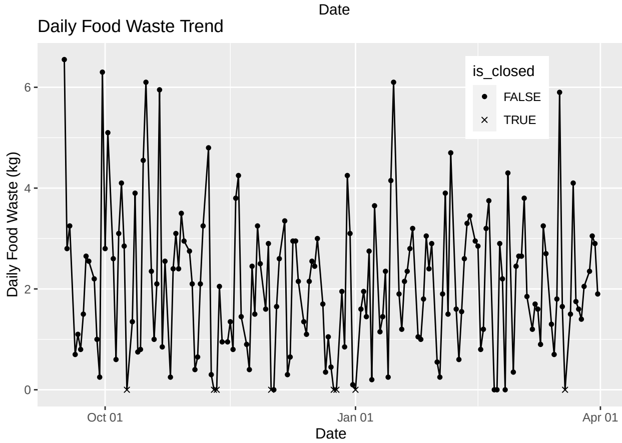


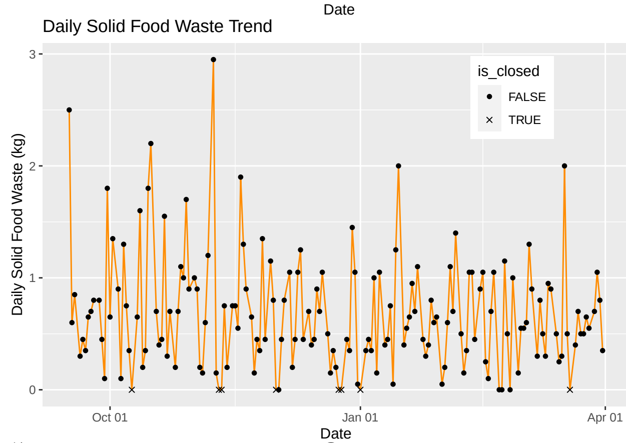
Time Series Plots —

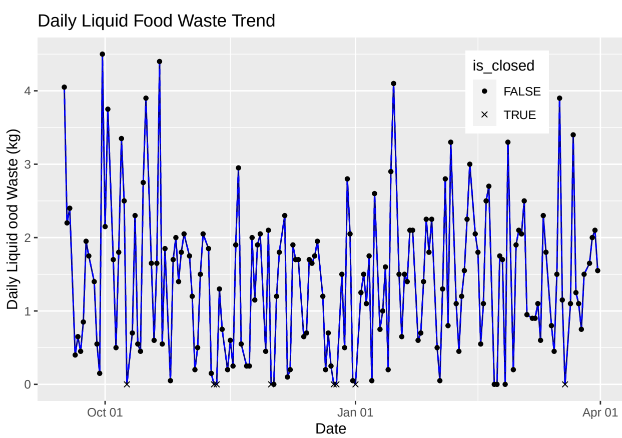
Daily Time Series

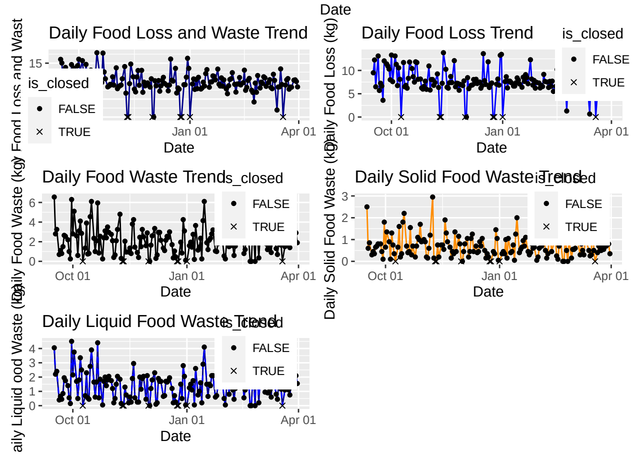












Decompsiotion

```
## -- Attaching packages --------------------------------------------- fpp3 0.5 --

## v tsibble     1.1.3     v fable       0.3.3
## v tsibbledata 0.4.1     v fabletools  0.3.4
## v feasts      0.3.1

## -- Conflicts -------------------------------------------------- fpp3_conflicts --
## x dplyr::combine()       masks gridExtra::combine()
## x lubridate::date()      masks base::date()
## x rstatix::filter()      masks dplyr::filter(), stats::filter()
## x tsibble::intersect()   masks base::intersect()
## x tsibble::interval()    masks lubridate::interval()
## x dplyr::lag()           masks stats::lag()
## x fabletools::model()    masks bayesforecast::model()
## x MASS::select()         masks rstatix::select(), dplyr::select()
## x tsibble::setdiff()     masks base::setdiff()
## x qqplotr::stat_qq_line() masks ggplot2::stat_qq_line()
## x tsibble::union()       masks base::union()

##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,0,2) with non-zero mean : 595.2761
##  ARIMA(0,0,0) with non-zero mean : 607.2775
##  ARIMA(1,0,0) with non-zero mean : 598.3493
##  ARIMA(0,0,1) with non-zero mean : 606.2906
##  ARIMA(0,0,0) with zero mean     : 795.7987
##  ARIMA(1,0,2) with non-zero mean : 593.7226
##  ARIMA(0,0,2) with non-zero mean : 603.5818
##  ARIMA(1,0,1) with non-zero mean : 598.3892
##  ARIMA(1,0,3) with non-zero mean : 594.7845
##  ARIMA(0,0,3) with non-zero mean : 602.7266
##  ARIMA(2,0,1) with non-zero mean : 593.1346
##  ARIMA(2,0,0) with non-zero mean : 593.03
##  ARIMA(3,0,0) with non-zero mean : 591.0829
##  ARIMA(4,0,0) with non-zero mean : 593.9004
##  ARIMA(3,0,1) with non-zero mean : 593.1032
##  ARIMA(4,0,1) with non-zero mean : 594.6705
##  ARIMA(3,0,0) with zero mean     : 655.5828
##
##  Now re-fitting the best model(s) without approximations...
##
##  ARIMA(3,0,0) with non-zero mean : 600.6932
##
##  Best model: ARIMA(3,0,0) with non-zero mean

## Series: df$food_waste_kg
## ARIMA(3,0,0) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ar3    mean
```

```
##       0.1053  -0.2083  -0.1262  2.0746
## s.e.  0.0788   0.0769   0.0786  0.0871
##
## sigma^2 = 1.97:  log likelihood = -295.16
## AIC=600.33   AICc=600.69   BIC=615.97
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,0,2) with non-zero mean : 242.2204
##  ARIMA(0,0,0) with non-zero mean : 254.9591
##  ARIMA(1,0,0) with non-zero mean : 242.9804
##  ARIMA(0,0,1) with non-zero mean : 254.9337
##  ARIMA(0,0,0) with zero mean     : 424.4576
##  ARIMA(1,0,2) with non-zero mean : 240.5345
##  ARIMA(0,0,2) with non-zero mean : 253.0456
##  ARIMA(1,0,1) with non-zero mean : 242.4608
##  ARIMA(1,0,3) with non-zero mean : 241.1252
##  ARIMA(0,0,3) with non-zero mean : 252.9766
##  ARIMA(2,0,1) with non-zero mean : 240.7382
##  ARIMA(2,0,3) with non-zero mean : 243.1306
##  ARIMA(1,0,2) with zero mean     : 290.294
##
##  Now re-fitting the best model(s) without approximations...
##
##  ARIMA(1,0,2) with non-zero mean : 252.8433
##
##  Best model: ARIMA(1,0,2) with non-zero mean
```

```
## Series: df$solid_waste_kg
## ARIMA(1,0,2) with non-zero mean
##
## Coefficients:
##          ar1      ma1      ma2    mean
##       0.3933  -0.3011  -0.2195  0.6723
## s.e.  0.2334   0.2269   0.0728  0.0303
##
## sigma^2 = 0.2516:  log likelihood = -121.24
## AIC=252.48   AICc=252.84   BIC=268.12
```

```
##
##  Fitting models using approximations to speed things up...
##
##  ARIMA(2,0,2) with non-zero mean : 481.848
##  ARIMA(0,0,0) with non-zero mean : 489.7931
##  ARIMA(1,0,0) with non-zero mean : 483.6428
##  ARIMA(0,0,1) with non-zero mean : 488.6056
##  ARIMA(0,0,0) with zero mean     : 668.5145
##  ARIMA(1,0,2) with non-zero mean : 481.4292
##  ARIMA(0,0,2) with non-zero mean : 487.558
##  ARIMA(1,0,1) with non-zero mean : 484.5832
##  ARIMA(1,0,3) with non-zero mean : 482.8695
##  ARIMA(0,0,3) with non-zero mean : 487.0004
```

```
##  ARIMA(2,0,1) with non-zero mean : 480.5155
##  ARIMA(2,0,0) with non-zero mean : 480.0232
##  ARIMA(3,0,0) with non-zero mean : 478.3711
##  ARIMA(4,0,0) with non-zero mean : 480.7297
##  ARIMA(3,0,1) with non-zero mean : 480.1401
##  ARIMA(4,0,1) with non-zero mean : 479.0072
##  ARIMA(3,0,0) with zero mean     : 539.5893
##
##  Now re-fitting the best model(s) without approximations...
##
##  ARIMA(3,0,0) with non-zero mean : 484.9027
##
##  Best model: ARIMA(3,0,0) with non-zero mean
```

```
## Series: df$liquid_waste_kg
## ARIMA(3,0,0) with non-zero mean
##
## Coefficients:
##          ar1      ar2     ar3    mean
##       0.1128  -0.1804  -0.124  1.4030
## s.e.  0.0780   0.0767   0.078  0.0638
##
## sigma^2 = 0.9932:  log likelihood = -237.27
## AIC=484.53   AICc=484.9   BIC=500.18
```

```
##
##  ARIMA(2,1,2) with drift         : Inf
##  ARIMA(0,1,0) with drift         : 382.2608
##  ARIMA(1,1,0) with drift         : 376.6995
##  ARIMA(0,1,1) with drift         : Inf
##  ARIMA(0,1,0)                    : 380.2918
##  ARIMA(2,1,0) with drift         : 371.6764
##  ARIMA(3,1,0) with drift         : 361.5494
##  ARIMA(4,1,0) with drift         : 358.102
##  ARIMA(5,1,0) with drift         : 360.2444
##  ARIMA(4,1,1) with drift         : Inf
##  ARIMA(3,1,1) with drift         : Inf
##  ARIMA(5,1,1) with drift         : Inf
##  ARIMA(4,1,0)                    : 355.9381
##  ARIMA(3,1,0)                    : 359.4474
##  ARIMA(5,1,0)                    : 358.0249
##  ARIMA(4,1,1)                    : 344.9549
##  ARIMA(3,1,1)                    : 342.6938
##  ARIMA(2,1,1)                    : 343.3855
##  ARIMA(3,1,2)                    : 344.9619
##  ARIMA(2,1,0)                    : 369.616
##  ARIMA(2,1,2)                    : 342.9415
##  ARIMA(4,1,2)                    : 347.2447
##
##  Best model: ARIMA(3,1,1)
```

```
## Series: df[1:92, ]$food_waste_kg
## ARIMA(3,1,1)
```

```
##
## Coefficients:
##          ar1      ar2      ar3      ma1
##       0.1433  -0.1843  -0.1961  -0.9352
## s.e.  0.1118   0.1076   0.1129   0.0380
##
## sigma^2 = 2.284:  log likelihood = -165.99
## AIC=341.99   AICc=342.69   BIC=354.54
```

```
##
##  ARIMA(2,0,2) with non-zero mean : 165.4809
##  ARIMA(0,0,0) with non-zero mean : 162.51
##  ARIMA(1,0,0) with non-zero mean : 163.1611
##  ARIMA(0,0,1) with non-zero mean : 162.7369
##  ARIMA(0,0,0) with zero mean     : 247.8297
##  ARIMA(1,0,1) with non-zero mean : 164.7709
##
##  Best model: ARIMA(0,0,0) with non-zero mean
```

```
## Series: df[1:92, ]$solid_waste_kg
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##         mean
##       0.7207
## s.e.  0.0597
##
## sigma^2 = 0.3311:  log likelihood = -79.19
## AIC=162.38   AICc=162.51   BIC=167.42
```

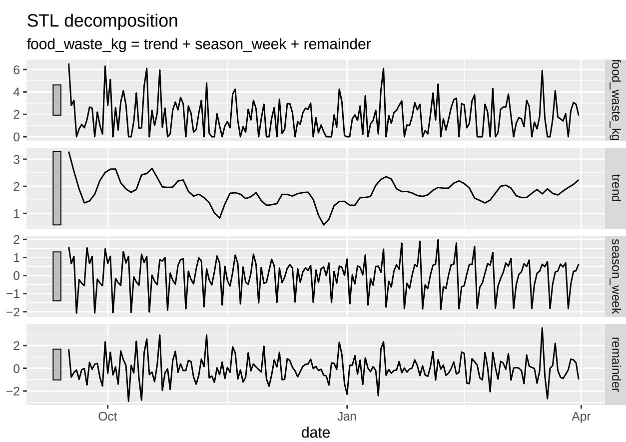
```
##
##  ARIMA(2,1,2) with drift         : Inf
##  ARIMA(0,1,0) with drift         : 315.6767
##  ARIMA(1,1,0) with drift         : 309.1532
##  ARIMA(0,1,1) with drift         : Inf
##  ARIMA(0,1,0)                    : 313.6831
##  ARIMA(2,1,0) with drift         : 303.7267
##  ARIMA(3,1,0) with drift         : 292.6036
##  ARIMA(4,1,0) with drift         : 287.7742
##  ARIMA(5,1,0) with drift         : 289.7147
##  ARIMA(4,1,1) with drift         : Inf
##  ARIMA(3,1,1) with drift         : Inf
##  ARIMA(5,1,1) with drift         : Inf
##  ARIMA(4,1,0)                    : 285.6019
##  ARIMA(3,1,0)                    : 290.4933
##  ARIMA(5,1,0)                    : 287.4865
##  ARIMA(4,1,1)                    : 278.0896
##  ARIMA(3,1,1)                    : 275.979
##  ARIMA(2,1,1)                    : 276.3443
##  ARIMA(3,1,2)                    : 278.1815
##  ARIMA(2,1,0)                    : 301.653
##  ARIMA(2,1,2)                    : 277.205
##  ARIMA(4,1,2)                    : 280.5544
```

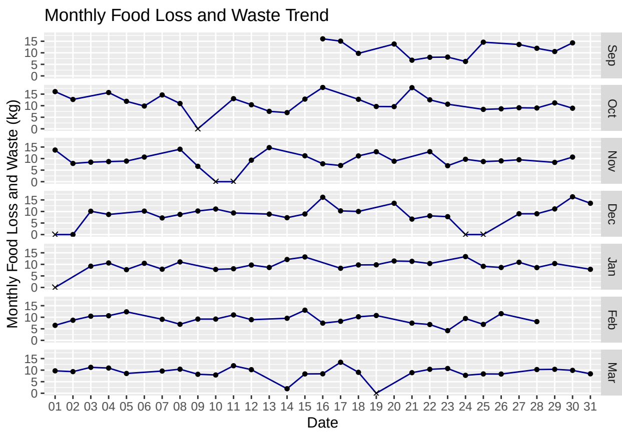
```
##
##  Best model: ARIMA(3,1,1)

## Series: df[1:92, ]$liquid_waste_kg
## ARIMA(3,1,1)
##
## Coefficients:
##          ar1      ar2      ar3      ma1
##       0.1304  -0.1809  -0.1865  -0.9185
## s.e.  0.1141   0.1076   0.1145   0.0510
##
## sigma^2 = 1.101:  log likelihood = -132.64
## AIC=275.27   AICc=275.98   BIC=287.83

##
##  ARIMA(2,0,2) with non-zero mean : Inf
##  ARIMA(0,0,0) with non-zero mean : 264.1095
##  ARIMA(1,0,0) with non-zero mean : 266.2064
##  ARIMA(0,0,1) with non-zero mean : 266.0714
##  ARIMA(0,0,0) with zero mean     : 360.2653
##  ARIMA(1,0,1) with non-zero mean : Inf
##
##  Best model: ARIMA(0,0,0) with non-zero mean

## Series: df[93:169, ]$food_waste_kg
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##         mean
##       2.1032
## s.e.  0.1491
##
## sigma^2 = 1.735:  log likelihood = -129.97
## AIC=263.95   AICc=264.11   BIC=268.63

##
##  ARIMA(2,0,2) with non-zero mean : 86.42921
##  ARIMA(0,0,0) with non-zero mean : 86.32735
##  ARIMA(1,0,0) with non-zero mean : 88.43897
##  ARIMA(0,0,1) with non-zero mean : 88.33825
##  ARIMA(0,0,0) with zero mean     : 174.9761
##  ARIMA(1,0,1) with non-zero mean : Inf
##
##  Best model: ARIMA(0,0,0) with non-zero mean

## Series: df[93:169, ]$solid_waste_kg
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##         mean
##       0.6188
## s.e.  0.0470
##
```

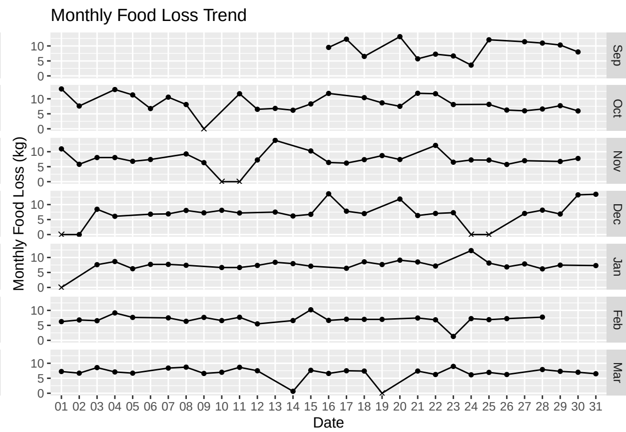
```
## sigma^2 = 0.1724:  log likelihood = -41.08
## AIC=86.17   AICc=86.33   BIC=90.85
```

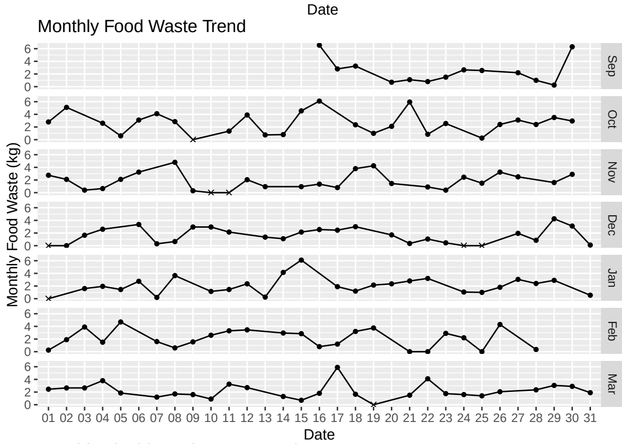
```
##
##  ARIMA(2,0,2) with non-zero mean : Inf
##  ARIMA(0,0,0) with non-zero mean : 214.2947
##  ARIMA(1,0,0) with non-zero mean : 216.4005
##  ARIMA(0,0,1) with non-zero mean : 216.3053
##  ARIMA(0,0,0) with zero mean     : 307.6959
##  ARIMA(1,0,1) with non-zero mean : Inf
##
##  Best model: ARIMA(0,0,0) with non-zero mean
```

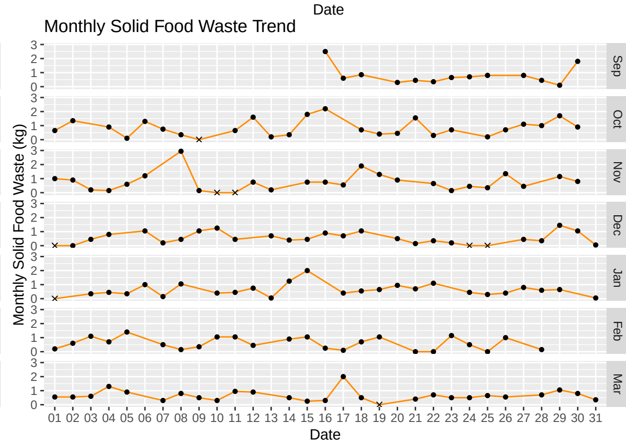
```
## Series: df[93:169, ]$liquid_waste_kg
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##         mean
##       1.4844
## s.e.  0.1079
##
## sigma^2 = 0.9086:  log likelihood = -105.07
## AIC=214.13   AICc=214.29   BIC=218.82
```

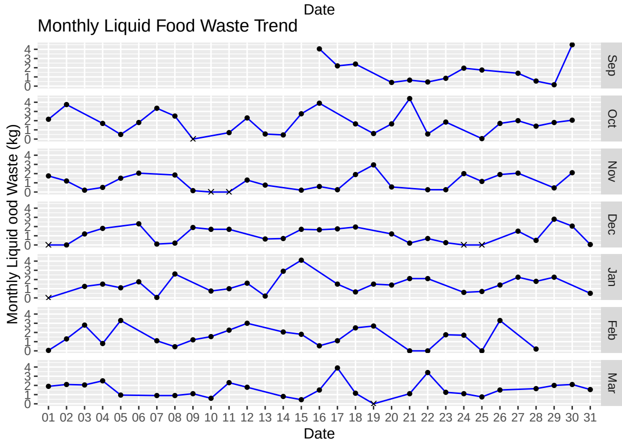



Monthly Food Loss and Waste Trend
Monthly Food Loss and Waste (kg)
Date
Sep
Oct
Nov
Dec
Jan
Feb
Mar

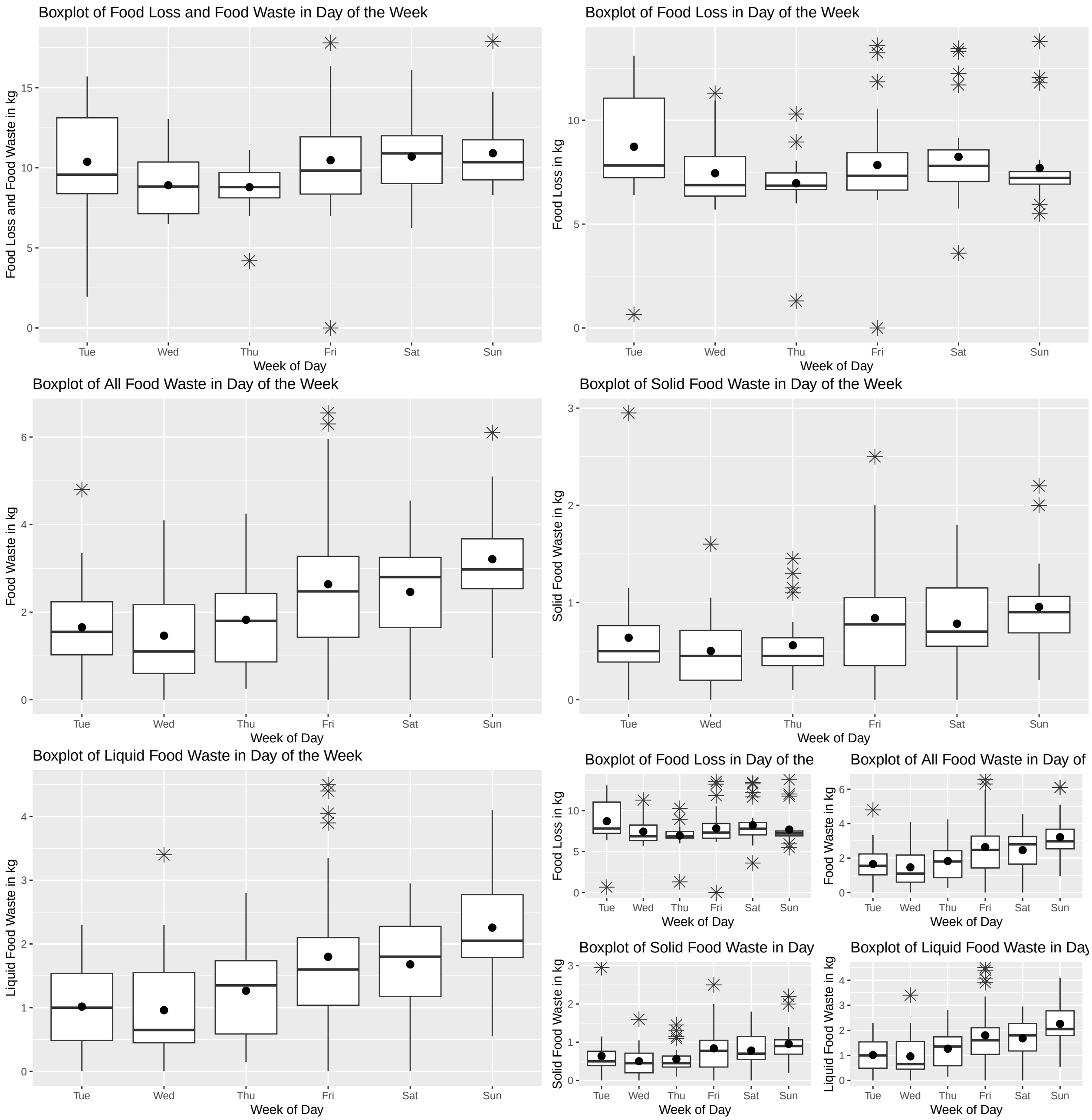

Monthly Food Loss Trend
Monthly Food Loss (kg)
Date
Sep
Oct
Nov
Dec
Jan
Feb
Mar


Monthly Food Waste Trend
Monthly Food Waste (kg)
Date
Sep
Oct
Nov
Dec
Jan
Feb
Mar

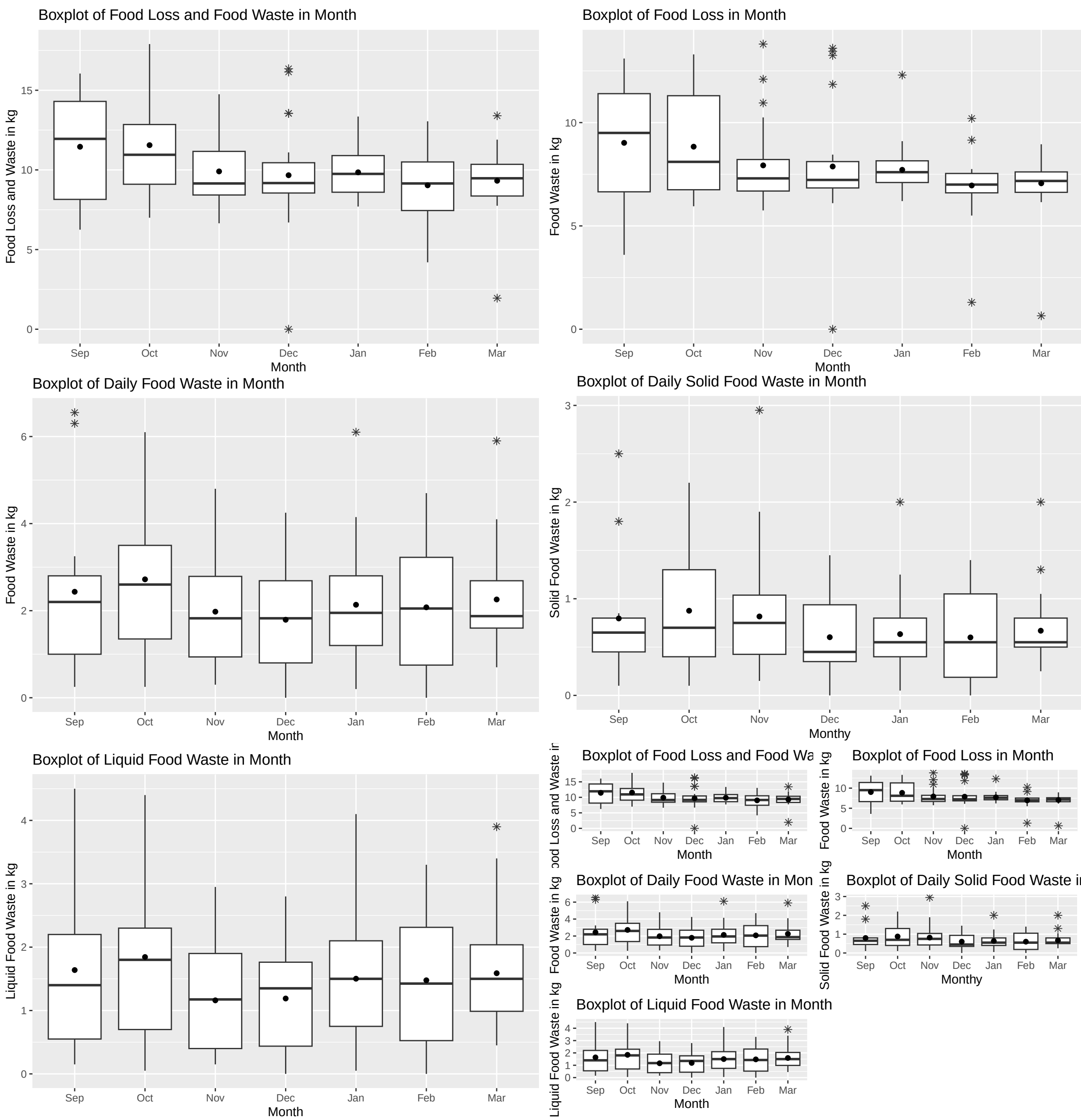

Monthly Solid Food Waste Trend
Monthly Solid Food Waste (kg)
Date
Sep
Oct
Nov
Dec
Jan
Feb
Mar


Monthly Liquid Food Waste Trend
Monthly Liquid ood Waste (kg)
Date
Sep
Oct
Nov
Dec
Jan
Feb
Mar

Boxplots

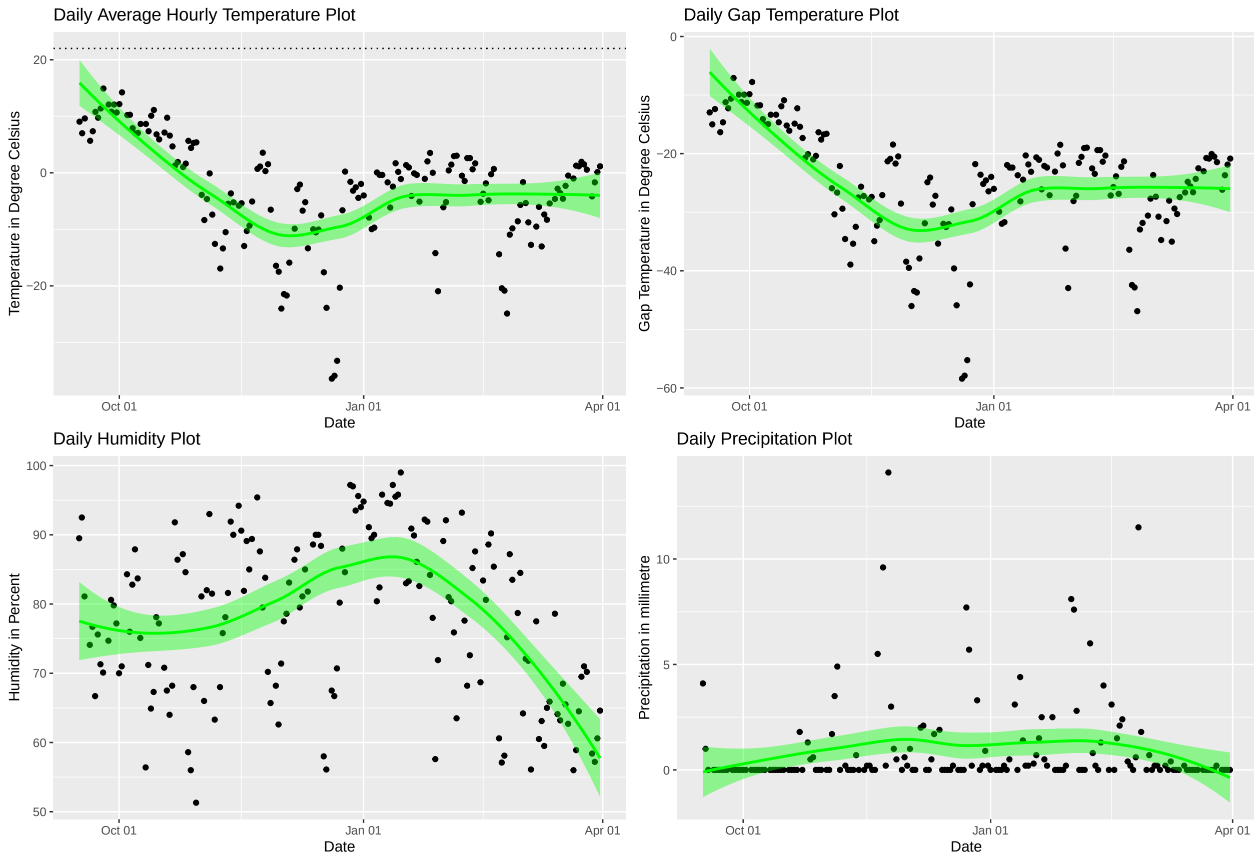


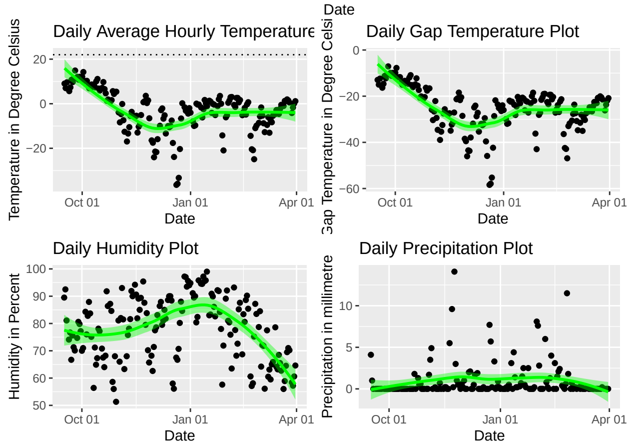
Boxplot of Food Loss and Food Waste in Day of the Week
Food Loss and Food Waste in kg
Week of Day
Boxplot of Food Loss in Day of the Week
Food Loss in kg
Week of Day
Boxplot of All Food Waste in Day of the Week
Food Waste in kg
Week of Day
Boxplot of Solid Food Waste in Day of the Week
Solid Food Waste in kg
Week of Day
Boxplot of Liquid Food Waste in Day of the Week
Liquid Food Waste in kg
Week of Day
Boxplot of Food Loss in Day of the
Boxplot of All Food Waste in Day of
Boxplot of Solid Food Waste in Day
Boxplot of Liquid Food Waste in Day
Tue
Wed
Thu
Fri
Sat
Sun



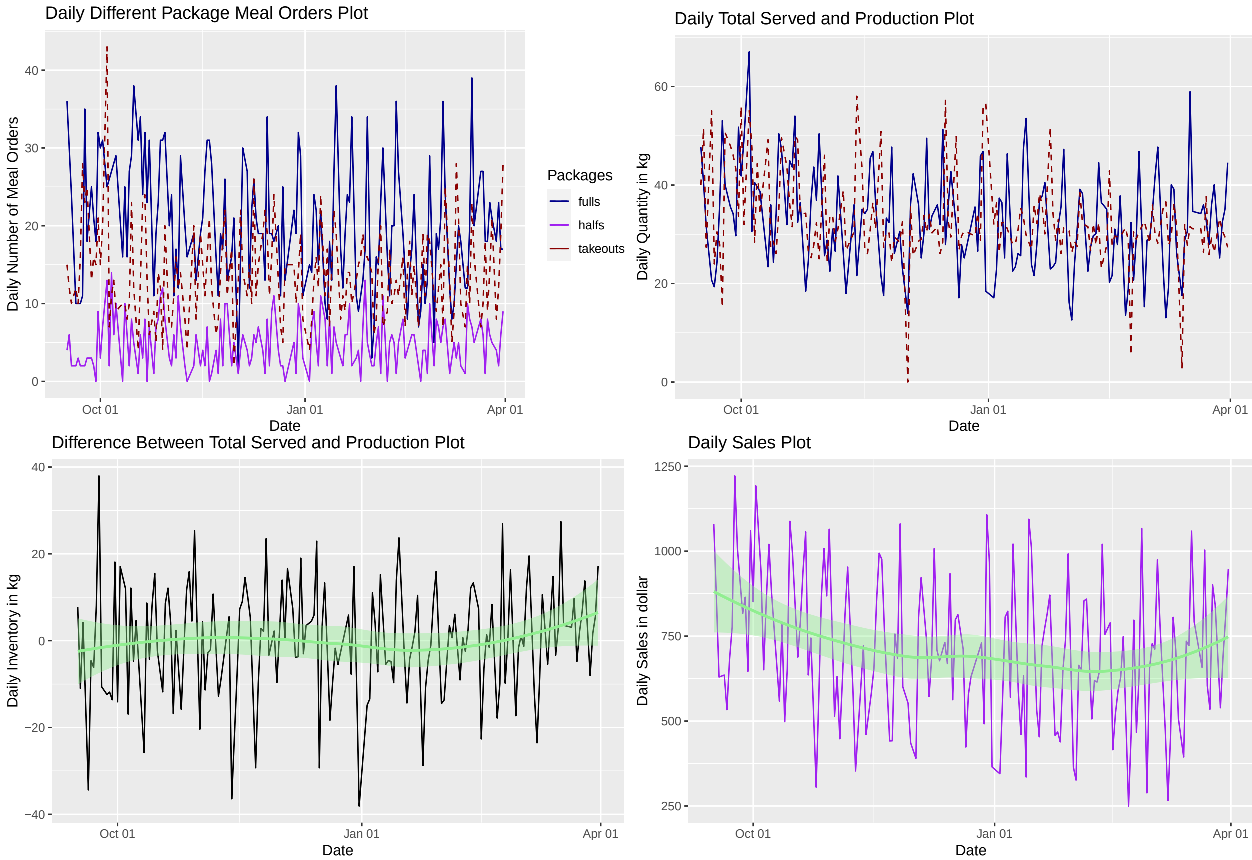
Boxplot of Food Loss and Food Waste in Month
Food Loss and Waste in kg
Month
Boxplot of Food Loss in Month
Food Waste in kg
Month
Boxplot of Daily Food Waste in Month
Food Waste in kg
Month
Boxplot of Daily Solid Food Waste in Month
Solid Food Waste in kg
Monthy
Boxplot of Liquid Food Waste in Month
Liquid Food Waste in kg
Month
Sep
Oct
Nov
Dec
Jan
Feb
Mar

Time Series Plots for Independents



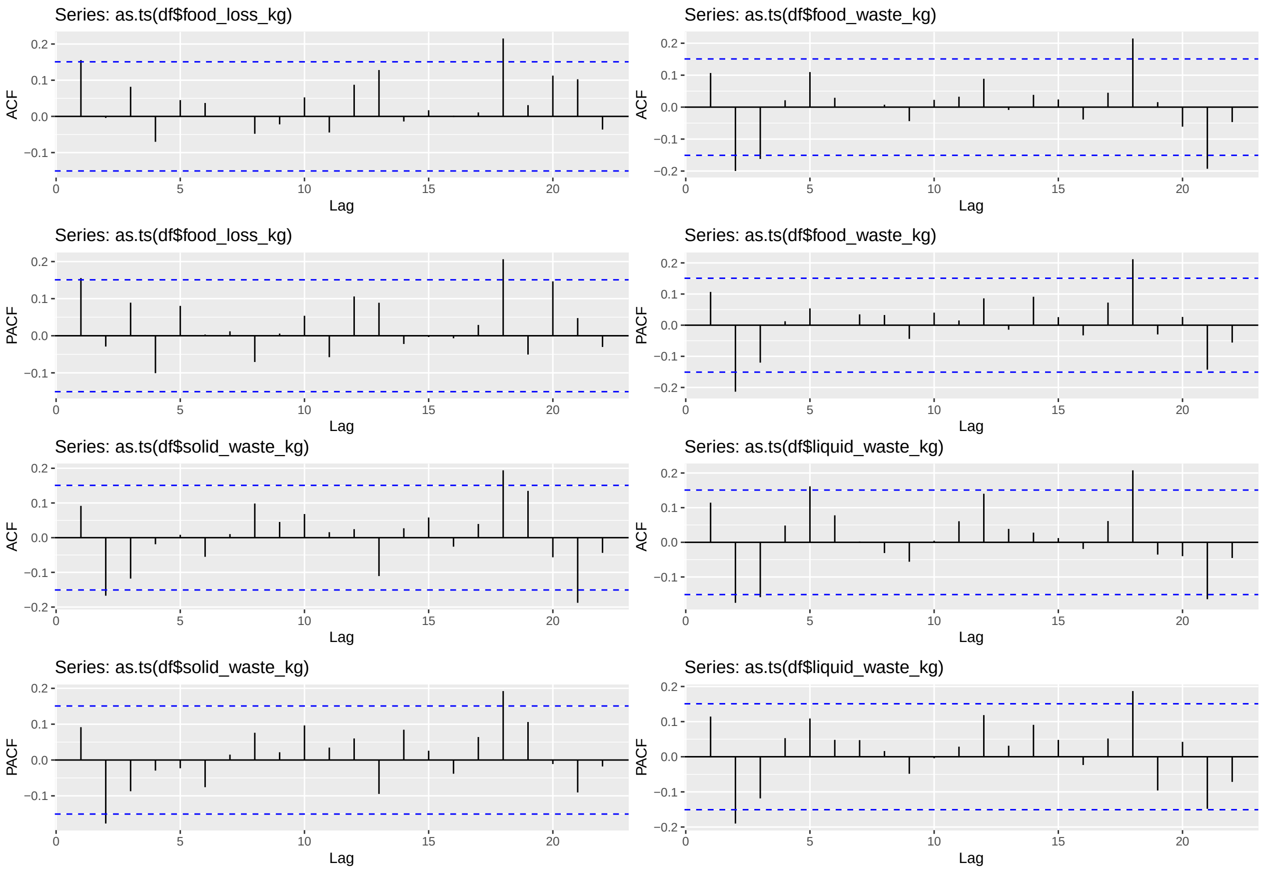



```
## 'geom_smooth()' using formula = 'y ~ x'
## 'geom_smooth()' using formula = 'y ~ x'
## 'geom_smooth()' using formula = 'y ~ x'
## 'geom_smooth()' using formula = 'y ~ x'
```


Daily Different Package Meal Orders Plot
Daily Number of Meal Orders
Date
Oct 01
Jan 01
Apr 01
Packages
fulls
halfs
takeouts
Daily Total Served and Production Plot
Daily Quantity in kg
Difference Between Total Served and Production Plot
Daily Inventory in kg
Daily Sales Plot
Daily Sales in dollar


Daily Different Package Meal Orde
Daily Number of Meal Orders
Packages
fulls
halfs
takeouts
Daily Total Served and Production
Daily Quantity in kg
Difference Between Total Served a
Daily Inventory in kg
Daily Sales Plot
Daily Sales in dollar
Date
Oct 01
Jan 01
Apr 01

(Partial and) Autocorrelation Function



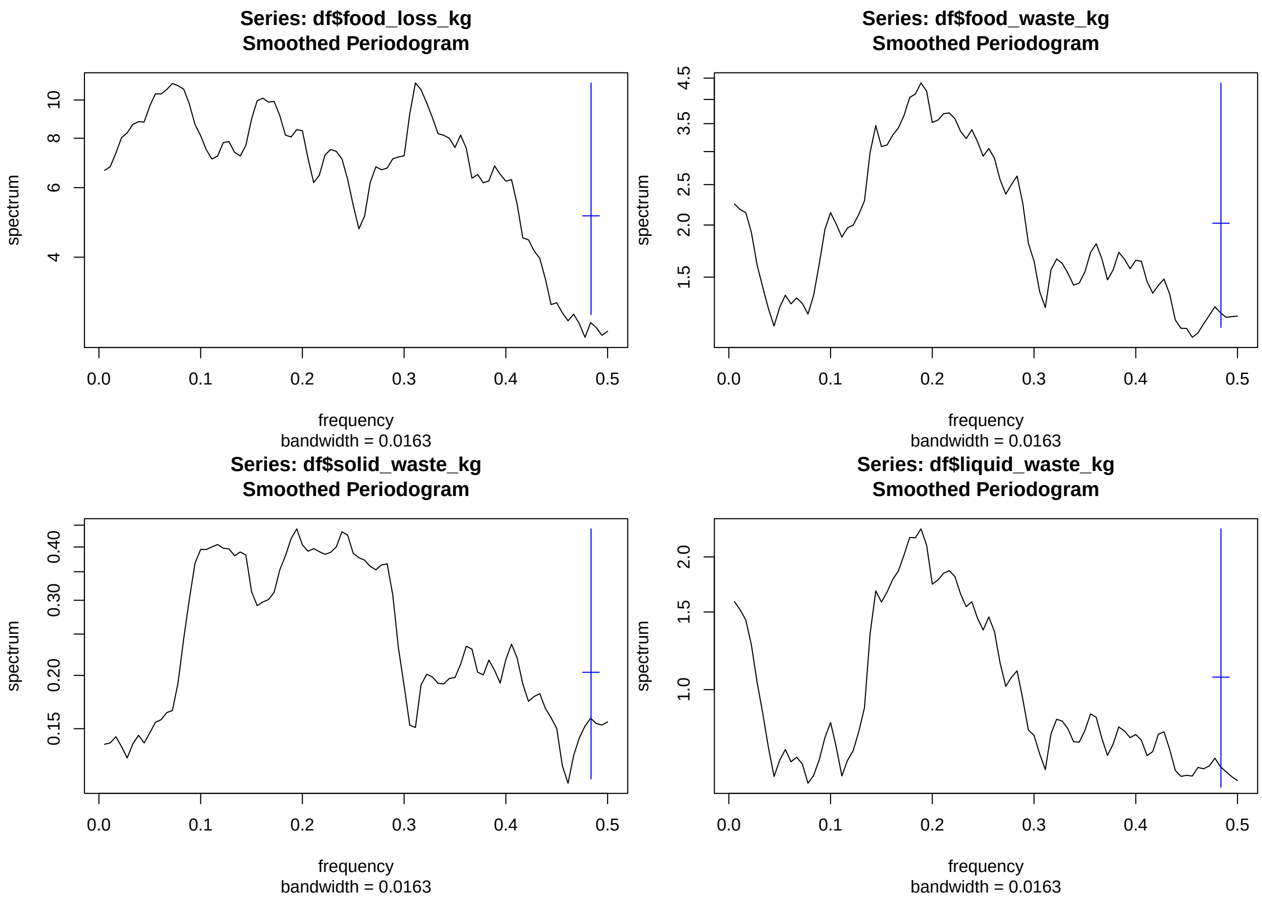
Spectral Analysis

```
## [1] 3.214286
```

```
## [1] 5.294118
```

```
## [1] 5.142857
```

```
## [1] 5.294118
```



roughly 6 (days) period for food waste, but food loss is approx. 3 days or 20 days cycle.