preliminary-results

Daisy Fang

4/14/2021

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v tibble  3.0.6     v purrr   0.3.4
## v tidyr   1.1.2     v dplyr   1.0.4
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x readr::guess_encoding()  masks rvest::guess_encoding()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x purrr::pluck()           masks rvest::pluck()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```
library(lubridate)
library(ggpubr)
library(broom)
library(purrr)
```

```
all_pallets <- read_csv("all_pallets.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   pallet_id = col_character(),
##   qty_per_mast = col_double(),
##   date_in = col_datetime(format = ""),
##   orig_qty = col_double(),
##   earliest_date = col_datetime(format = ""),
##   latest_date = col_datetime(format = ""),
##   pick_count = col_double(),
##   qty_picked = col_double(),
##   room_type = col_character(),
##   item_desc = col_character(),
##   product_group = col_character(),
##   stackable = col_character(),
##   pallet_weight = col_double(),
##   blast_hours = col_double()
## )
```

```
all_pallets <- all_pallets %>%
  mutate(item_category = case_when(str_detect(item_desc, "MASHED|MASH|MSHD|SR CRM & CHIVE POTAS") ~ "MAS
                                   str_detect(item_desc, "POT|WEDGE") ~ "OTHER POTATO",
                                   str_detect(item_desc, "MACARONI|MAC") ~ "MAC & CHEESE",
                                   str_detect(item_desc, "LINKS|LINK|LIN|PROLL|ROLL|PATTIES|PATTIS|PAT|
                                   TRUE ~ "OTHERS"))

pallets_dates <- all_pallets %>%
  #filter(item_category == "MASHED POTATOES") %>%
  filter(orig_qty == qty_picked) %>%
  select(item_category, date_in, latest_date) %>%
  mutate(date_in = date(date_in),
         date_out = date(latest_date)) %>%
  mutate(week = week(date_out),
         month = month(date_out),
         year = as.factor(year(date_out))) %>%
  select(-latest_date)
```

```
potatoes <- pallets_dates %>%
  filter(item_category == "MASHED POTATOES") %>%
  filter(week < 53)
```

```
mashedpotato <- all_pallets %>%
  filter(item_category == "MASHED POTATOES") %>%
  filter(orig_qty == qty_picked) %>%
  select(item_category, date_in, latest_date) %>%
  mutate(date_in = date(date_in),
         #year_week_in = paste(year(date_in), " W", week(date_in), sep = ""),
         date_out = date(latest_date),
         #year_week_out = paste(year(date_out), " W", week(date_out), sep = "")
         ) %>%
  select(-latest_date) %>%
  group_by(date_out) %>%
  count() %>%
  mutate(week = week(date_out),
         month = month(date_out),
         year = as.factor(year(date_out))) %>%
  filter(week != 53)
```

COVID Panic Buying

```
jan19 <- mashedpotato %>%
  filter(month == 1, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
jan20 <- mashedpotato %>%
  filter(month == 1, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
t1 <- t.test(jan19$n, jan20$n, paired = TRUE)
```

```
feb19 <- mashedpotato %>%
  filter(month == 2, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
feb20 <- mashedpotato %>%
  filter(month == 2, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
t2 <- t.test(feb19$n, feb20$n[1:28], paired = TRUE)

mar19 <- mashedpotato %>%
  filter(month == 3, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
mar20 <- mashedpotato %>%
  filter(month == 3, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
t3 <- t.test(mar19$n, mar20$n, paired = TRUE)

apr19 <- mashedpotato %>%
  filter(month == 4, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
apr20 <- mashedpotato %>%
  filter(month == 4, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
t.test(apr19$n, apr20$n, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  apr19$n and apr20$n
## t = 0.47949, df = 29, p-value = 0.6352
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -33.74259  54.40926
## sample estimates:
## mean of the differences
##                10.33333
```

```
may19 <- mashedpotato %>%
  filter(month == 5, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
may20 <- mashedpotato %>%
  filter(month == 5, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
t.test(may19$n, may20$n, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  may19$n and may20$n
## t = -3.16, df = 30, p-value = 0.003589
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -169.08914  -36.33022
## sample estimates:
## mean of the differences
##                -102.7097
```

```r
jun19 <- mashedpotato %>%
  filter(month == 6, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
jun20 <- mashedpotato %>%
  filter(month == 6, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
t.test(jun19$n, jun20$n, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  jun19$n and jun20$n
## t = -1.1732, df = 29, p-value = 0.2503
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -89.98068  24.38068
## sample estimates:
## mean of the differences
##                   -32.8
```

```r
jul19 <- mashedpotato %>%
  filter(month == 7, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
jul20 <- mashedpotato %>%
  filter(month == 7, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
t.test(jul19$n, jul20$n, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  jul19$n and jul20$n
## t = -1.2987, df = 30, p-value = 0.2039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -82.40340  18.33888
## sample estimates:
## mean of the differences
##               -32.03226
```

```r
aug19 <- mashedpotato %>%
  filter(month == 8, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
aug20 <- mashedpotato %>%
  filter(month == 8, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
t.test(aug19$n, aug20$n, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  aug19$n and aug20$n
## t = 0.43914, df = 30, p-value = 0.6637
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -46.98642  72.72835
## sample estimates:
## mean of the differences
##                12.87097
```

```r
sep19 <- mashedpotato %>%
  filter(month == 9, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
sep20 <- mashedpotato %>%
  filter(month == 9, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```r
t.test(sep19$n, sep20$n, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  sep19$n and sep20$n
```

```
## t = 1.7687, df = 29, p-value = 0.08747
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.796661 93.729994
## sample estimates:
## mean of the differences
##                43.46667
```

```
oct19 <- mashedpotato %>%
  filter(month == 10, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
oct20 <- mashedpotato %>%
  filter(month == 10, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
t.test(oct19$n, oct20$n, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  oct19$n and oct20$n
## t = 0.76041, df = 30, p-value = 0.4529
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -39.91444  87.26928
## sample estimates:
## mean of the differences
##                23.67742
```

```
nov19 <- mashedpotato %>%
  filter(month == 11, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
nov20 <- mashedpotato %>%
  filter(month == 11, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
t.test(nov19$n[1:29], nov20$n, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  nov19$n[1:29] and nov20$n
## t = 0.87894, df = 28, p-value = 0.3869
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -48.31275 120.93344
## sample estimates:
## mean of the differences
```

```
##                36.31034
```

```
dec19 <- mashedpotato %>%
  filter(month == 12, year == 2019) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
dec20 <- mashedpotato %>%
  filter(month == 12, year == 2020) %>%
  select(n, year)
```

```
## Adding missing grouping variables: `date_out`
```

```
t.test(dec19$n[1:28], dec20$n, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  dec19$n[1:28] and dec20$n
## t = 2.0754, df = 27, p-value = 0.0476
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##    0.7819195 136.9323662
## sample estimates:
## mean of the differences
##                68.85714
```

```
tab <- map_df(list(t1, t2, t3), tidy) %>%
  rename(t_statistic = statistic,
         p_value = p.value,
         degree_of_freedom = parameter) %>%
  mutate(conf_int = paste(round(conf.low, 2), round(conf.high,2), sep = ", ")) %>%
  mutate(conf_int = paste("(", conf_int, ")")) %>%
  select(t_statistic, degree_of_freedom, p_value, conf_int)

tab
```

```
## # A tibble: 3 x 4
##   t_statistic degree_of_freedom p_value conf_int
##         <dbl>             <dbl>   <dbl> <chr>
## 1       -3.21                29 0.00327 ( -90.8, -20.07 )
## 2       -1.51                27 0.144   ( -77.65, 11.93 )
## 3       -3.17                30 0.00348 ( -185.58, -40.22 )
```

```
kable(tab, digits = 3)
```

t_statistic	degree_of_freedom	p_value	conf_int
-3.206	29	0.003	(-90.8, -20.07)
-1.505	27	0.144	(-77.65, 11.93)
-3.173	30	0.003	(-185.58, -40.22)

```
# Wilcox test
wilcox.test(n19, n20, paired = TRUE)
```

Serial Correlation

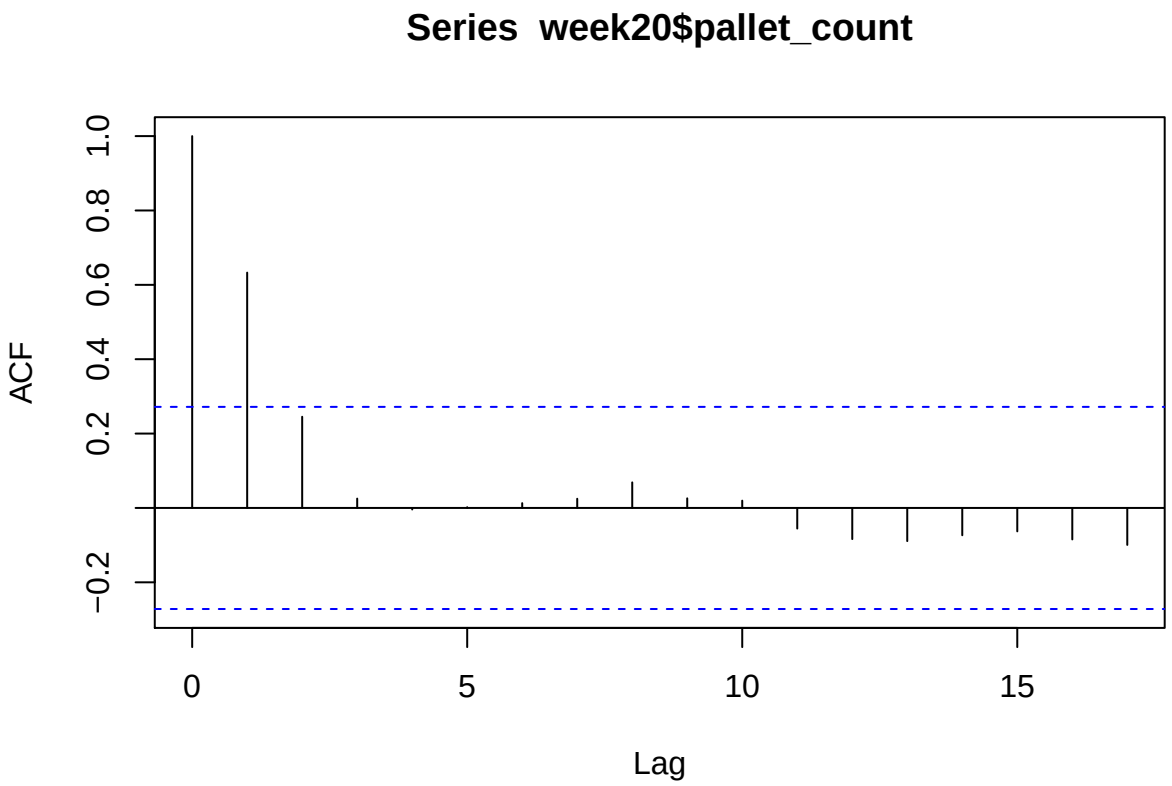
```
week20 <- potatoes %>%
  group_by(year, week) %>%
  summarize(pallet_count = n()) %>%
  filter(year == 2020)
```

```
## `summarise()` has grouped output by 'year'. You can override using the `.groups` argument.
```

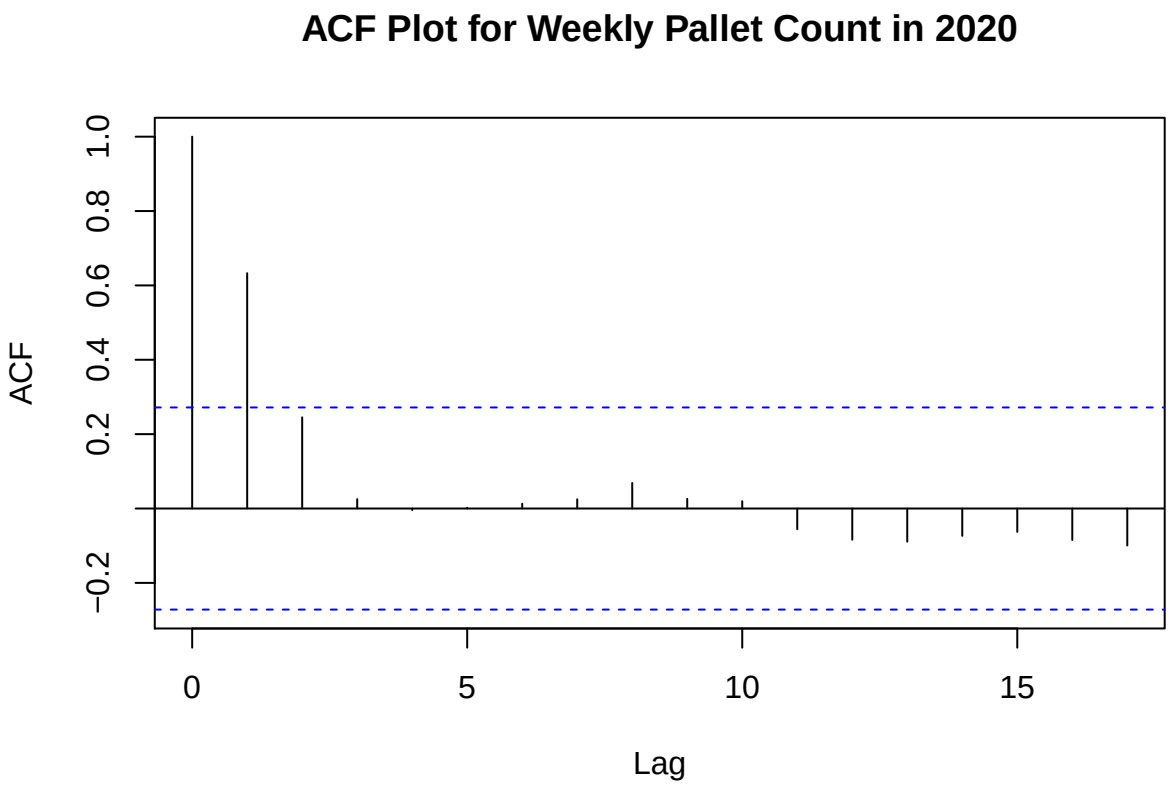
```
day20 <- potatoes %>%
  group_by(year, date_out) %>%
  summarize(pallet_count = n()) %>%
  filter(year == 2020)
```

```
## `summarise()` has grouped output by 'year'. You can override using the `.groups` argument.
```

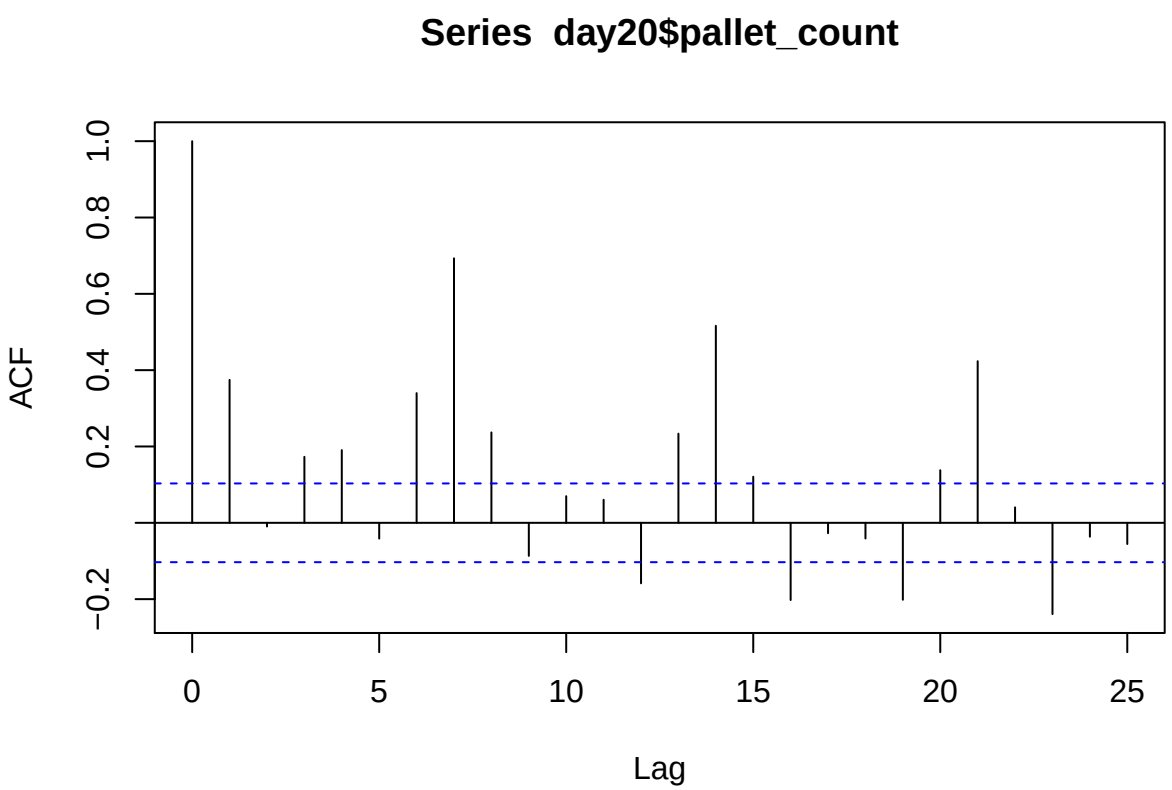
```
week20acf <- acf(week20$pallet_count)
```



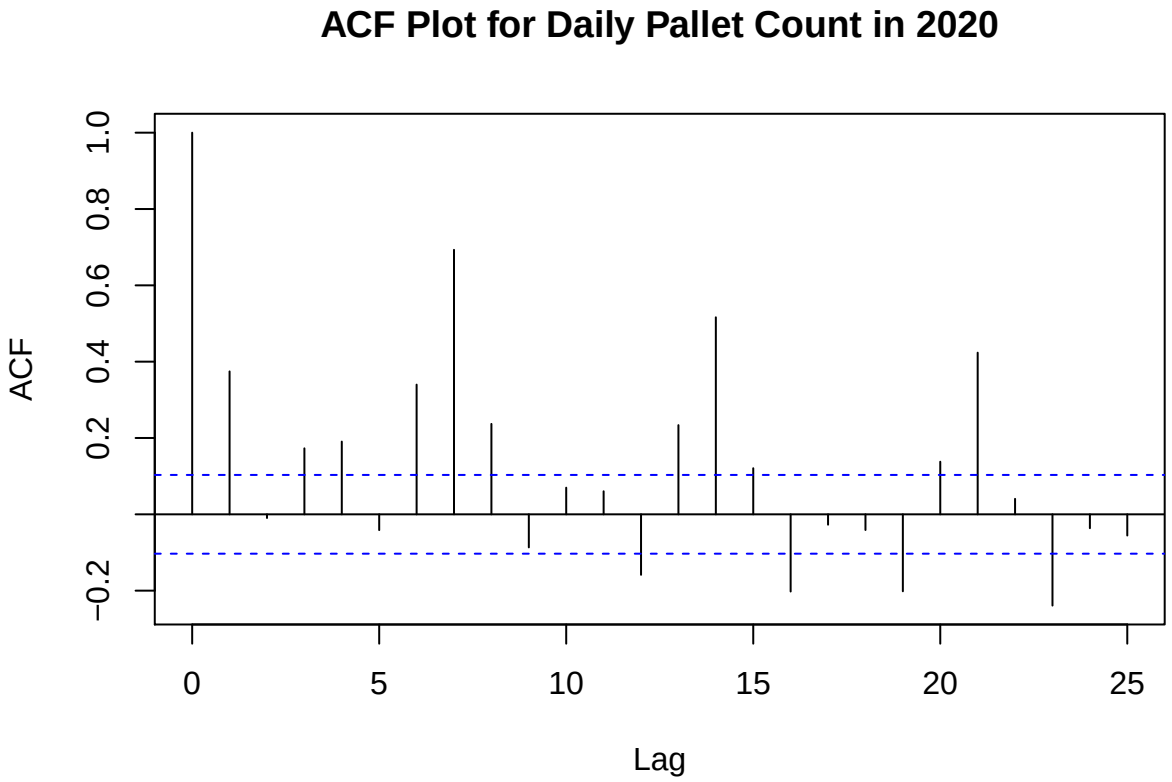

```
plot(week20acf, main = "ACF Plot for Weekly Pallet Count in 2020")
```




```
day20acf <- acf(day20$pallet_count)
```




```
plot(day20acf, main = "ACF Plot for Daily Pallet Count in 2020")
```



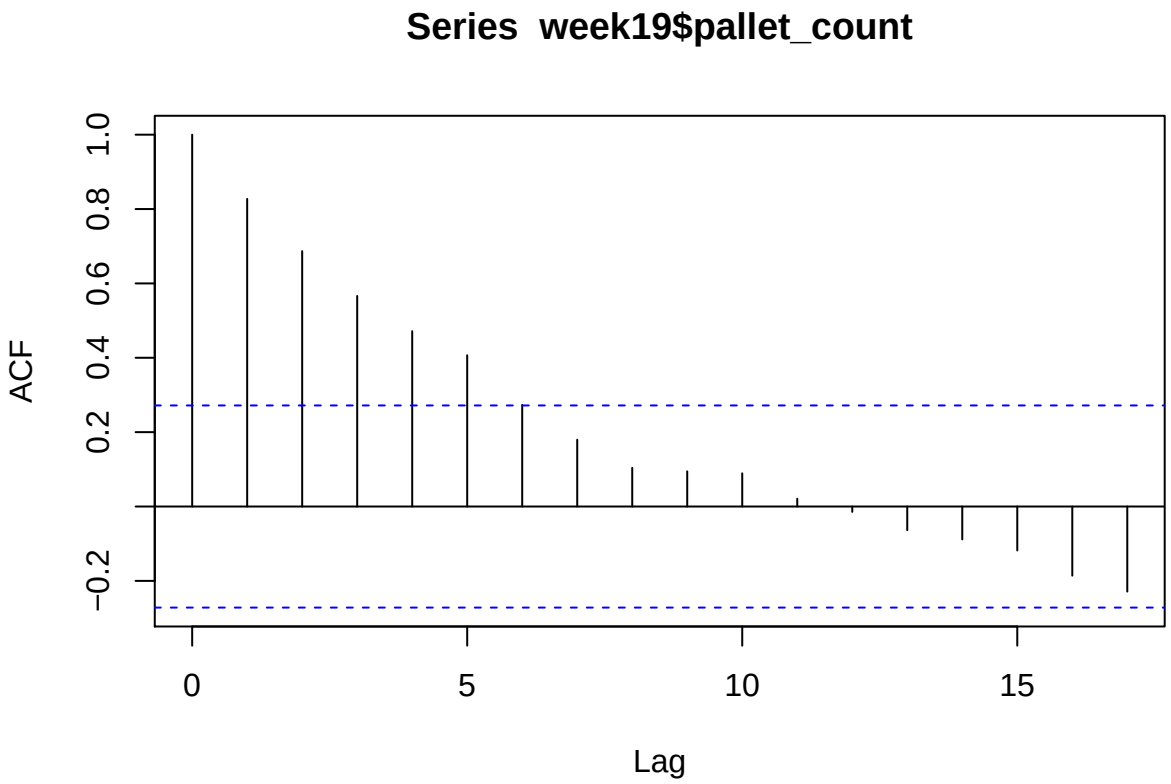

```
week19 <- potatoes %>%
  group_by(year, week) %>%
  summarize(pallet_count = n()) %>%
  filter(year == 2019)
```

```
## `summarise()` has grouped output by 'year'. You can override using the `.groups` argument.
```

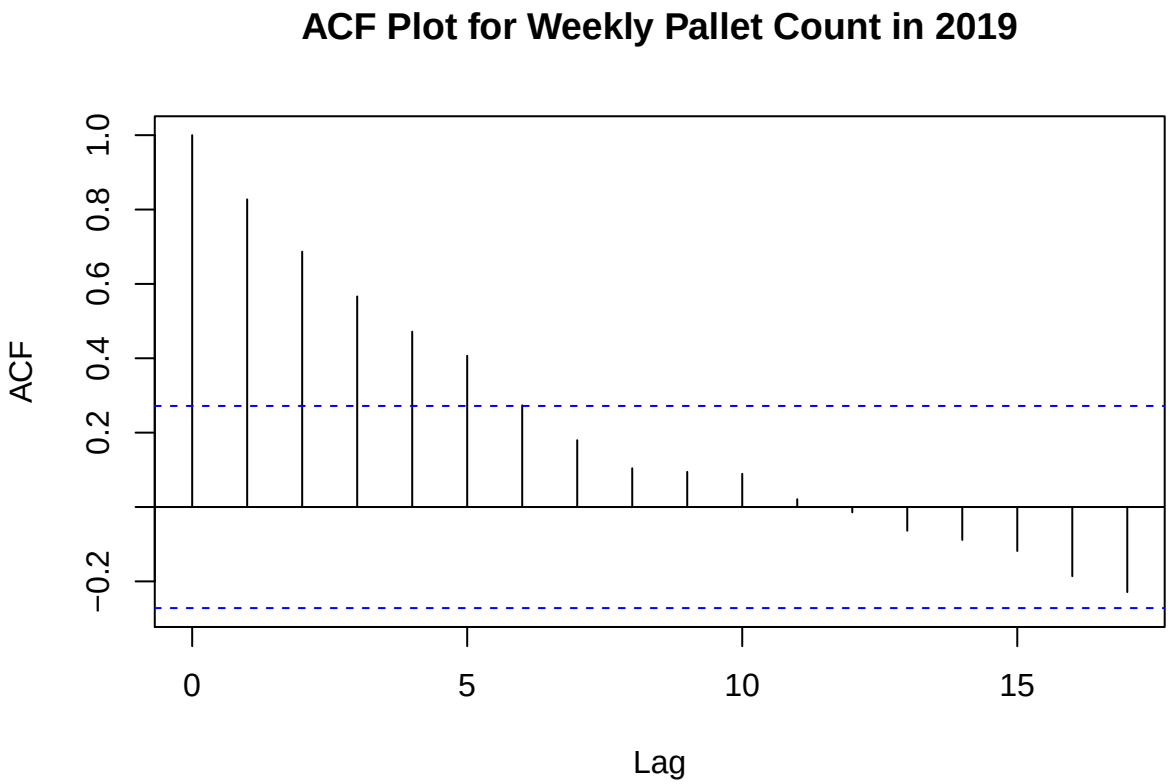
```
day19 <- potatoes %>%
  group_by(year, date_out) %>%
  summarize(pallet_count = n()) %>%
  filter(year == 2019)
```

```
## `summarise()` has grouped output by 'year'. You can override using the `.groups` argument.
```

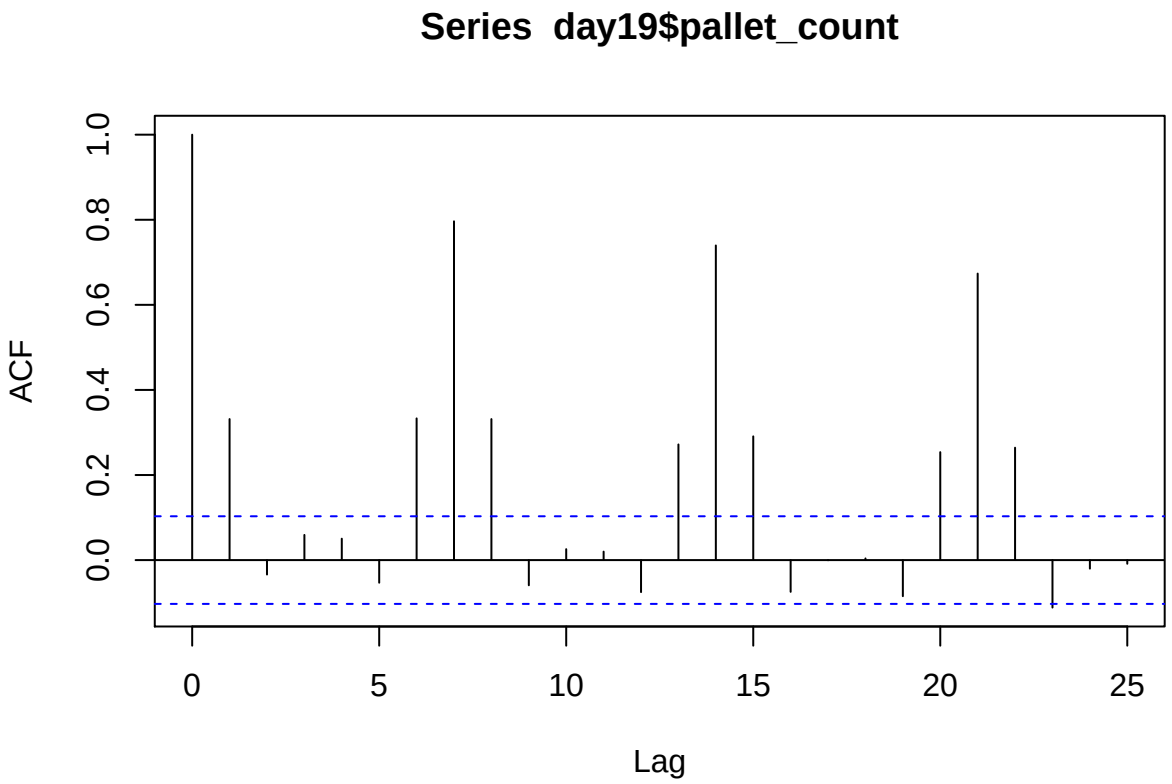
```
week19acf <- acf(week19$pallet_count)
```




```
plot(week19acf, main = "ACF Plot for Weekly Pallet Count in 2019")
```




```
day19acf <- acf(day19$pallet_count)
```




```
plot(day19acf, main = "ACF Plot for Daily Pallet Count in 2019")
```