Phoenix market analysis

```
# Load packages
library(tidyverse)
library(plyr)
library(dplyr)
library(tsibble)
library(ggplot2)
library(feasts)
library(lubridate)
```

```
# read data with phoenix -> chicago reefer lane, trucking volume, and yuma weather
data_raw <- readr::read_csv(file = "data/data_phoenix_with_yuma_weather_and_volume_and_lags.csv")  %>%
  mutate(yw = yearweek(yw)) %>%
  select(-X1) %>%
  as_tsibble(key = c(Mode,ORegionDAT, DRegionDAT), index = yw) %>%
  relocate(yw, Mode, ORegionDAT, DRegionDAT, approx_cost, approx_vol, tmax, prcp)
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   yw = col_character(),
##   Mode = col_character(),
##   ORegionDAT = col_character(),
##   DRegionDAT = col_character()
## )
## i Use `spec()` for the full column specifications.
```

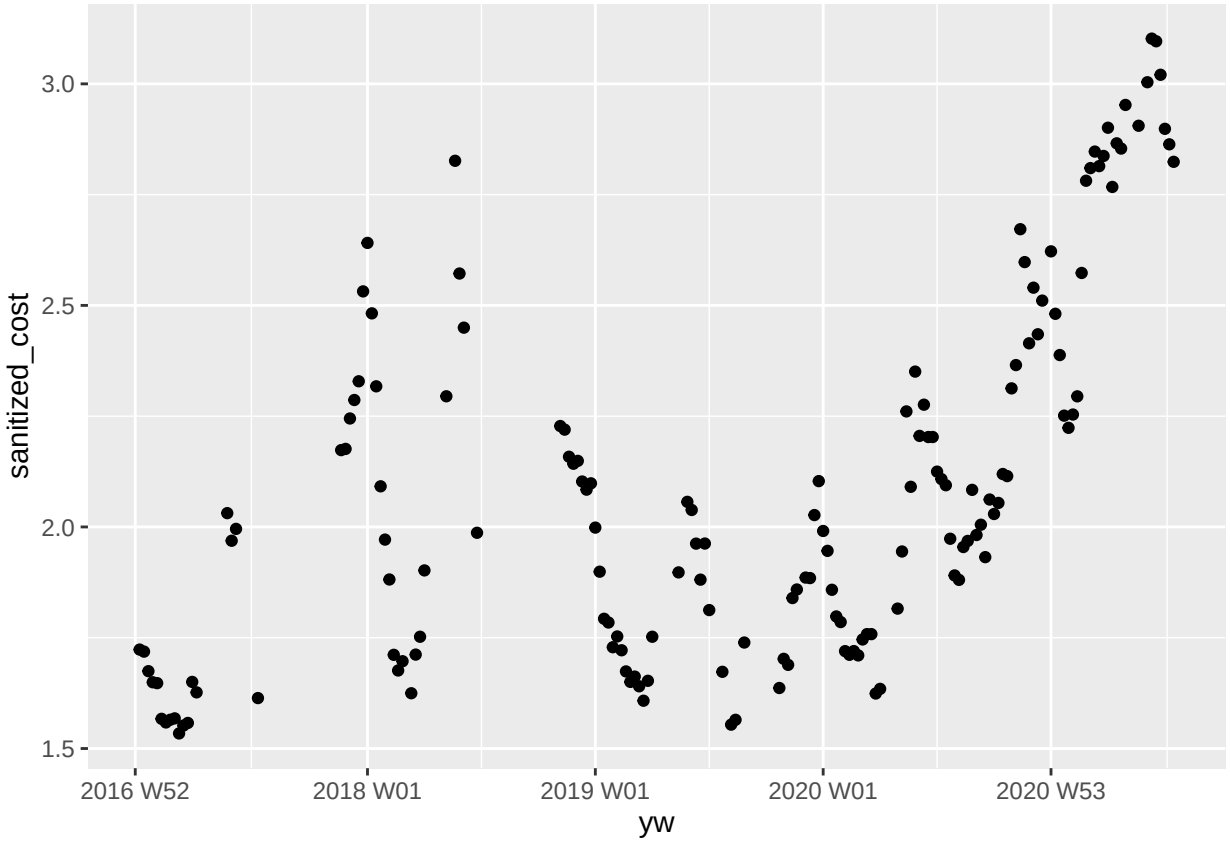
```
data_raw
```

```
## # A tsibble: 237 x 21 [1W]
## # Key:       Mode, ORegionDAT, DRegionDAT [1]
##          yw Mode  ORegionDAT DRegionDAT approx_cost approx_vol  tmax    prcp
##       <week> <chr> <chr>      <chr>            <dbl>      <dbl> <dbl>   <dbl>
##  1 2017 W01 R     AZ_PHO     IL_CHI            1.72       16.4  64.1 0
##  2 2017 W02 R     AZ_PHO     IL_CHI            1.72       13.1  68.1 0
##  3 2017 W03 R     AZ_PHO     IL_CHI            1.67       11.3  66   0.0257
##  4 2017 W04 R     AZ_PHO     IL_CHI            1.65       16.1  63.1 0
##  5 2017 W05 R     AZ_PHO     IL_CHI            1.65       13.7  74.9 0
##  6 2017 W06 R     AZ_PHO     IL_CHI            1.57       11.1  77.4 0.00286
##  7 2017 W07 R     AZ_PHO     IL_CHI            1.56       11.1  74.1 0.124
##  8 2017 W08 R     AZ_PHO     IL_CHI            1.56       14.9  69.7 0
##  9 2017 W09 R     AZ_PHO     IL_CHI            1.57       14.3  71.9 0.0471
## 10 2017 W10 R     AZ_PHO     IL_CHI            1.53       16.9  83.1 0
## # ... with 227 more rows, and 13 more variables: sanitized_cost <dbl>,
## #   sanitized_vol <dbl>, tmax_lag_12 <dbl>, tmax_lag_8 <dbl>, tmax_lag_4 <dbl>,
## #   tmax_lag_2 <dbl>, prcp_lag_12 <dbl>, prcp_lag_8 <dbl>, prcp_lag_4 <dbl>,
## #   prcp_lag_2 <dbl>, cluster_1 <dbl>, cluster_2 <dbl>, cluster_3 <dbl>
```

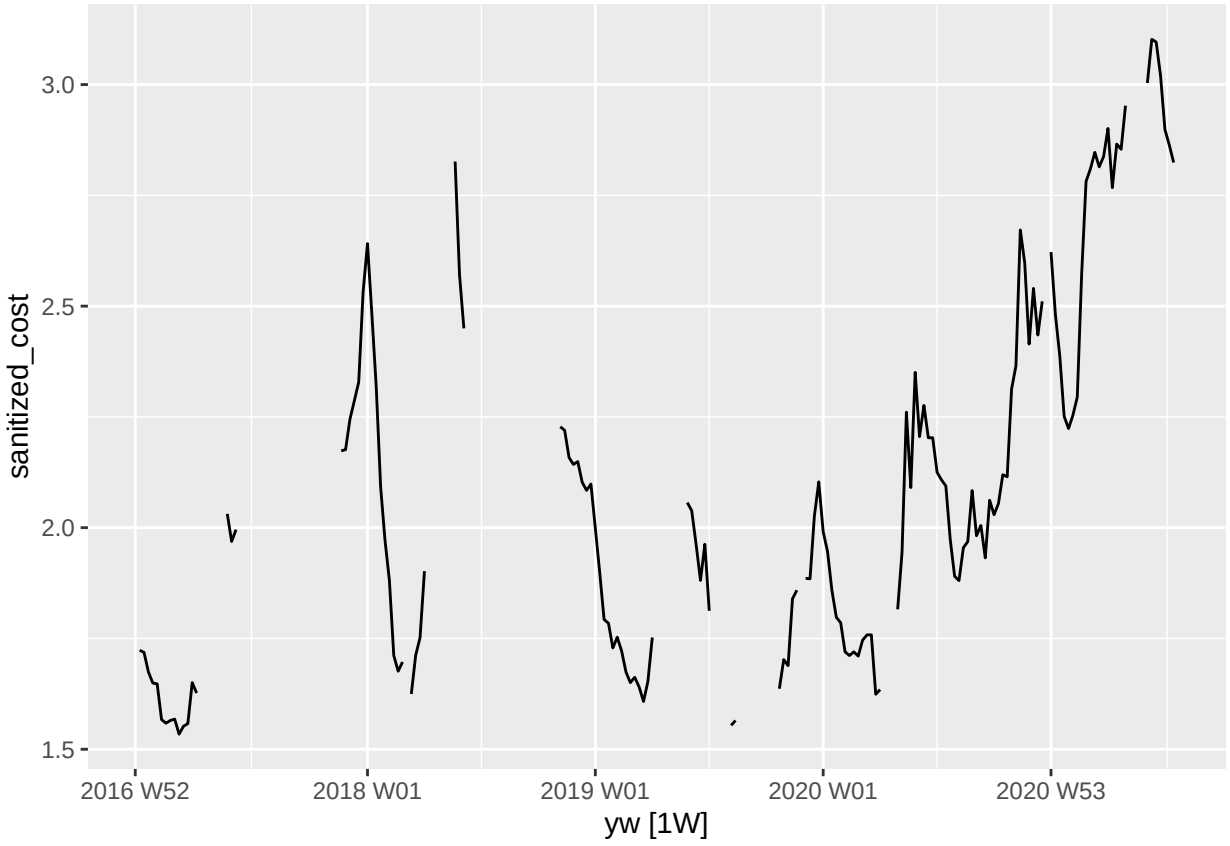
Plot the cost time series

```
#plot actual data values (some are missing)
ggplot(data_raw, aes(x=yw, y=sanitized_cost))+
  geom_point()
```

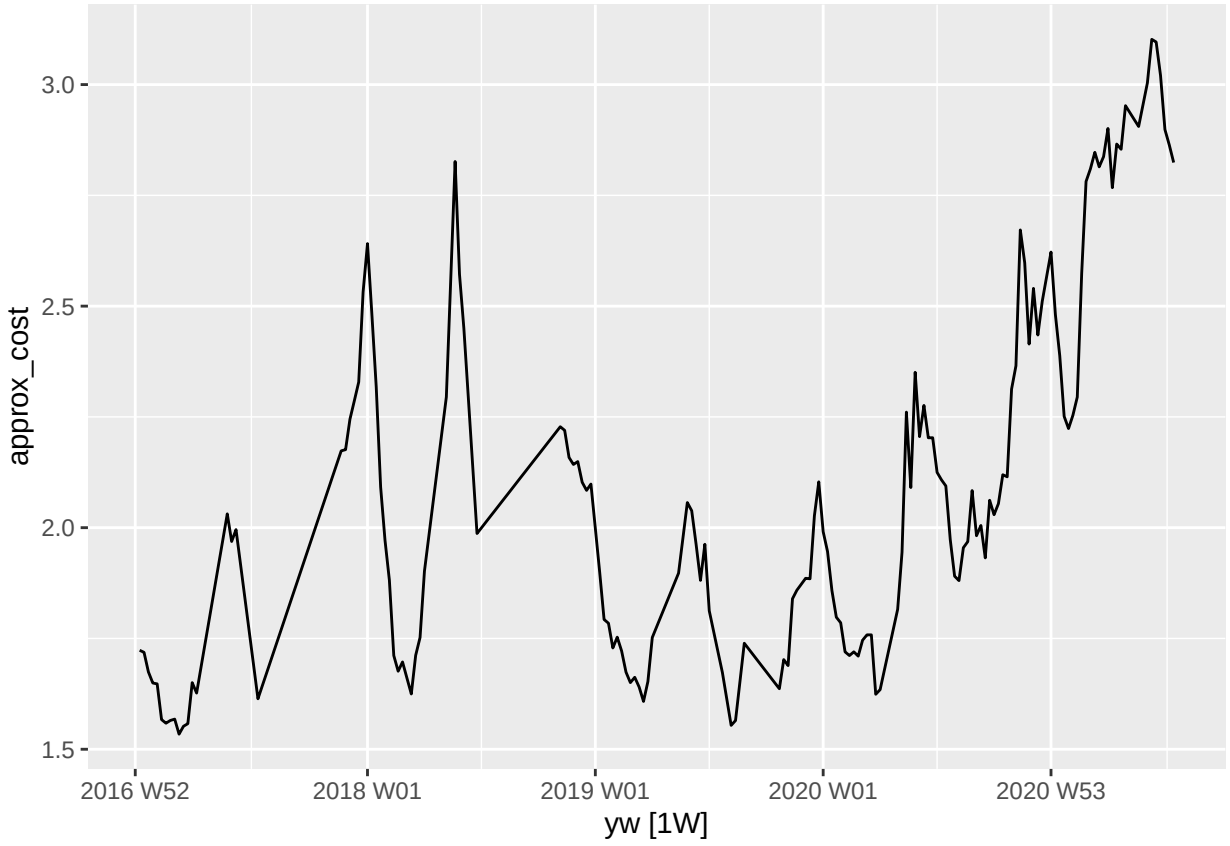
```
## Warning: Removed 79 rows containing missing values (geom_point).
```



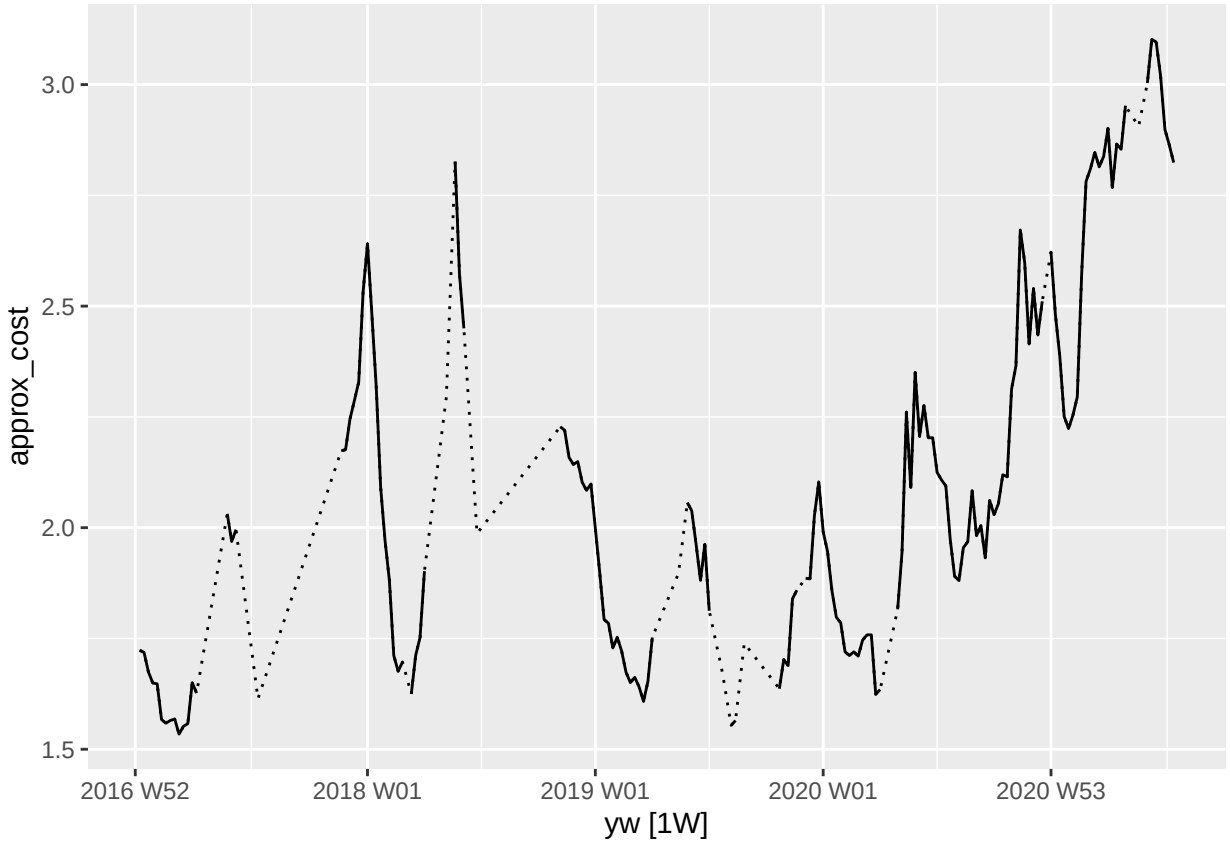

```
#line plot version
autoplot(data_raw, sanitized_cost)
```




```
#plot linearly interpolated data
autoplot(data_raw, approx_cost)
```

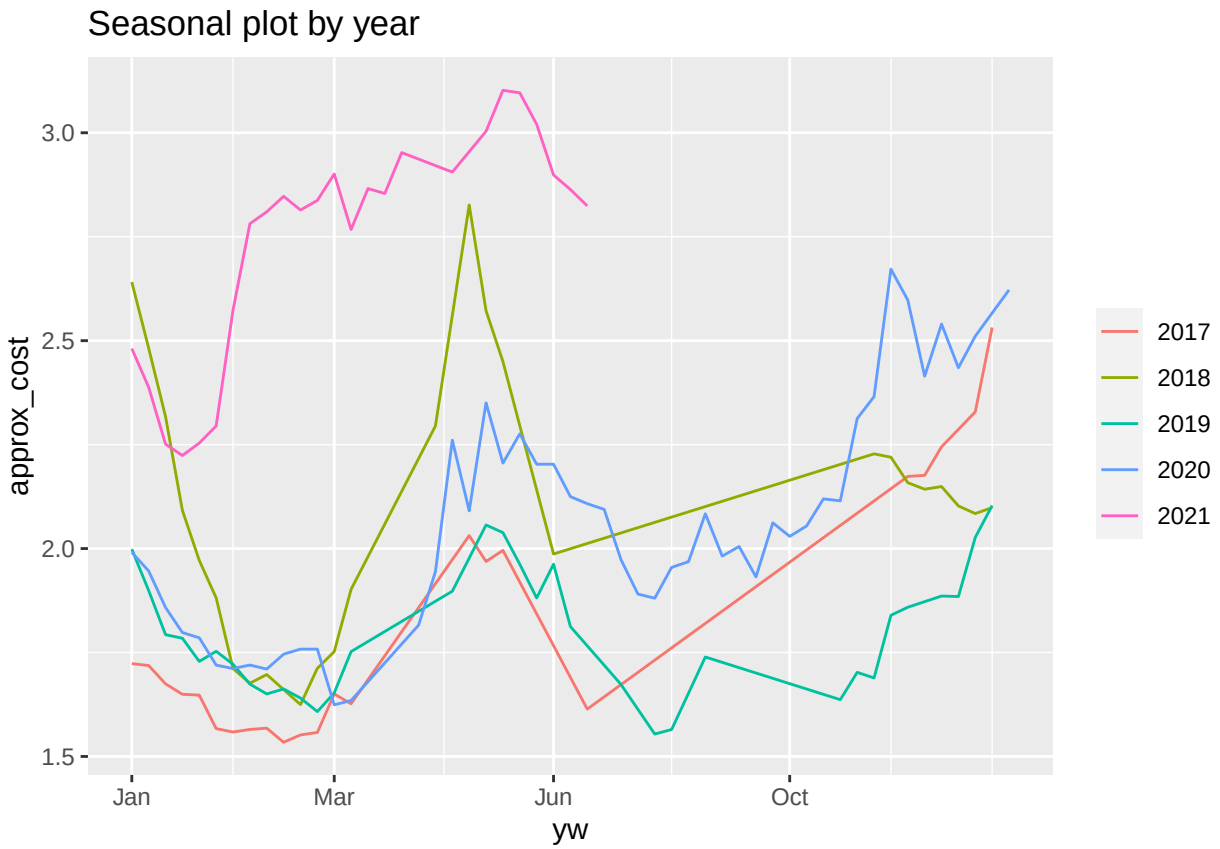



```
autoplot(data_raw, approx_cost, color="black", linetype="dotted") + autolayer(data_raw, sanitized_cost,
```



Seasonal plot

```
gg_season(data_raw, approx_cost) +
labs(title = "Seasonal plot by year")
```



I want to plot the cost, precip, temp, and volume data all together to see if I notice any visual patterns. I'll normalize all the values first so that the plot looks alright

Normalize

```
scaled_data <- data_raw %>%
  mutate_at(c("sanitized_cost", "approx_cost", "approx_vol", "prcp", "tmax"), ~(scale(.) %>% as.vector))

head(scaled_data)
```

```
## # A tsibble: 6 x 21 [1W]
## # Key:       Mode, ORegionDAT, DRegionDAT [1]
##         yw Mode  ORegionDAT DRegionDAT approx_cost approx_vol   tmax   prcp
##     <week> <chr> <chr>      <chr>            <dbl>      <dbl>  <dbl>  <dbl>
## 1 2017 W01 R     AZ_PHO     IL_CHI          -0.859      0.861 -1.58  -0.264
## 2 2017 W02 R     AZ_PHO     IL_CHI          -0.871      0.202 -1.33  -0.264
## 3 2017 W03 R     AZ_PHO     IL_CHI          -0.989     -0.140 -1.46   0.531
## 4 2017 W04 R     AZ_PHO     IL_CHI          -1.06       0.789 -1.65  -0.264
## 5 2017 W05 R     AZ_PHO     IL_CHI          -1.06       0.328 -0.896 -0.264
## 6 2017 W06 R     AZ_PHO     IL_CHI          -1.28      -0.183 -0.730 -0.175
## # ... with 13 more variables: sanitized_cost <dbl>, sanitized_vol <dbl>,
## #   tmax_lag_12 <dbl>, tmax_lag_8 <dbl>, tmax_lag_4 <dbl>, tmax_lag_2 <dbl>,
## #   prcp_lag_12 <dbl>, prcp_lag_8 <dbl>, prcp_lag_4 <dbl>, prcp_lag_2 <dbl>,
## #   cluster_1 <dbl>, cluster_2 <dbl>, cluster_3 <dbl>
```

Plot normalized shipping cost (black/grey), precip (blue), tmax (red), and shipping volume (purple).

```
scaled_data %>%
ggplot(aes(x=yw)) +
  geom_line(aes(y = tmax), color = "red") +
  geom_line(aes(y = prcp), color = "blue") +
  #geom_line(aes(y = sanitized_cost), color = "black") +
  geom_line(aes(y = approx_cost), color = "black") +
  geom_line(aes(y = approx_vol), color = "purple") +
  labs(title = "Phoenix -> Chicago Reefer", y= "normalized value")
```