census-permits-msa

suzanne

2023-08-11

In this task we need to download Excel-based and text based monthly Census reports of residential permit activity by MSA, and assembling the data for our two MSAs (Seattle-Tacoma-Bellevue and Bremerton-Silverdale-Port Orchard) into a consolidated monthly time-series dataset.

Prior to Nov-2019, the files are text based, then switch to excel.

```
library(readxl)
library(httr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(psrcplot)
library(ggplot2)
library(openxlsx)
library(stringr)
install_psrc_fonts()
```

Get the data from January 2019 to November 2019 from text files

A bunch of specific unchanging parts of the file and url:

```
col_names<-c("CSA", "CBSA", "Name", "Total", "1 Unit", "2 Units",
             "3 and 4 Units", "5 Units or More",
             "Num of Structures With 5 Units or More", "Monthly Coverage Percent*", "Month")

url_start<-'https://www.census.gov/construction/bps/txt/tb3u2019'
url_end<-'.txt'
month_ids<-c('01', '02', '03', '04', '05', '06', '07', '08', '09', '10')
num_cols<-c("CSA", "CBSA","Total", "1 Unit", "2 Units",
             "3 and 4 Units", "5 Units or More",
             "Num of Structures With 5 Units or More", "Monthly Coverage Percent*")
```

Get the data and clean it up

```
download_clean_text<-function(url, month_id,colnames){
    big_tbl<-read.fwf(url, skip=11, header=FALSE, widths=c(3,6,40,8,8,8,8,8,8,8))

    # Select the region's MSA; the Bremerton one is
    #spread across two lines and needs to be cleaned up
    psrc_msa_tbl<-big_tbl[c(47, 48,321),]

    col_names_no_month<-head(col_names, -1)


    psrc_msa_tbl<-setNames(psrc_msa_tbl, col_names_no_month)
    #clean up the Bremerton MSA data
    psrc_msa_tbl[1, 4:10]=psrc_msa_tbl[2,4:10]
    psrc_msa_tbl_clean<-psrc_msa_tbl[-2,]
    psrc_msa_tbl_clean$Month=paste0(2019,month_id)
return(psrc_msa_tbl_clean)
}
```

put the months together

```
df_2019<-data.frame(matrix(nrow = 0, ncol = length(col_names)))
colnames(df_2019)<-col_names

for(month_id in month_ids){
  url<-paste0(url_start, month_id, url_end)
  the_month_df<-download_clean_text(url, month_id,colnames)
  df_2019<-rbind(df_2019,the_month_df)
}
```

```
# for being consistent with later datasets
df_2019<-df_2019%>%mutate_at(num_cols,as.numeric)%>%
  select(-c("Monthly Coverage Percent*"))
```

#Part two: Get data from November 2019 to June 2023

```
month_ids<-c('01', '02', '03', '04', '05', '06', '07', '08', '09', '10', '11', '12')
yr_ids<-c('2020', '2021', '2022')

months_2019<-c('11','12')
months_2023<-c('01', '02', '03', '04', '05', '06')

month_yr_ids<-as.list(outer(yr_ids, month_ids, paste0))
months_2019_ids<-as.list(outer('2019', months_2019, paste0))
months_2023_ids<-as.list(outer('2023', months_2023, paste0))
month_yr_ids<-c(months_2019_ids,month_yr_ids,months_2023_ids)
url_start<-'https://www.census.gov/construction/bps/xls/msamonthly_'
url_end<-'.xls'
```

```
download_clean_excel<-function(url, month_id, months_19_ids){

    GET(url, write_disk(tf<-tempfile(fileext=".xls")))
```

```
  # sheet names changed after 2019
    if(month_id %in% months_2019_ids){
      sheetName='Units'}
    else{
      sheetName='MSA Units'
    }

    big_tbl<-read_excel(tf, sheet=sheetName,range='A8:I377')

    psrc_msa_tbl_clean<-big_tbl%>%filter(Name %in% c('Seattle-Tacoma-Bellevue, WA', 'Bremerton-Silverdal
    psrc_msa_tbl_clean$Month=month_id
return(psrc_msa_tbl_clean)
}
```

```
col_names<-col_names[names(col_names)!="Monthly Coverage Percent*"]
df_new<-data.frame(matrix(nrow = 0, ncol = length(col_names)))
colnames(df_new)<-col_names


for(month_id in month_yr_ids){
  url<-paste0(url_start, month_id, url_end)
  print(url)
  the_month_df<-download_clean_excel(url, month_id, months_19_ids)
  df_new<-rbind(df_new,the_month_df)
}
```

```
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_201911.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_201912.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202001.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202101.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202201.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202002.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202102.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202202.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202003.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202103.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202203.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202004.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202104.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202204.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202005.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202105.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202205.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202006.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202106.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202206.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202007.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202107.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202207.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202008.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202108.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202208.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202009.xls"
```

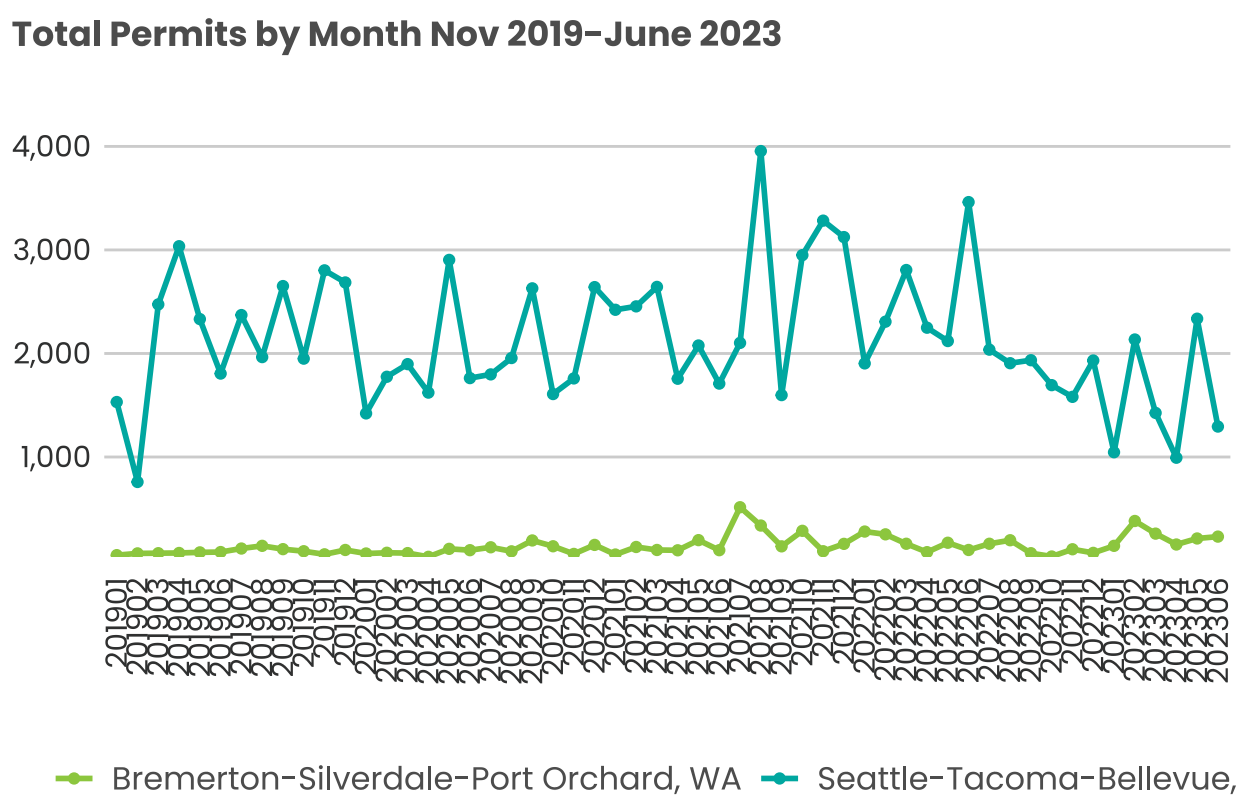
```
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202109.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202209.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202010.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202110.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202210.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202011.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202111.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202211.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202012.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202112.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202212.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202301.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202302.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202303.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202304.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202305.xls"
## [1] "https://www.census.gov/construction/bps/xls/msamonthly_202306.xls"
```

```
df_all<-rbind(df_2019, df_new)
```

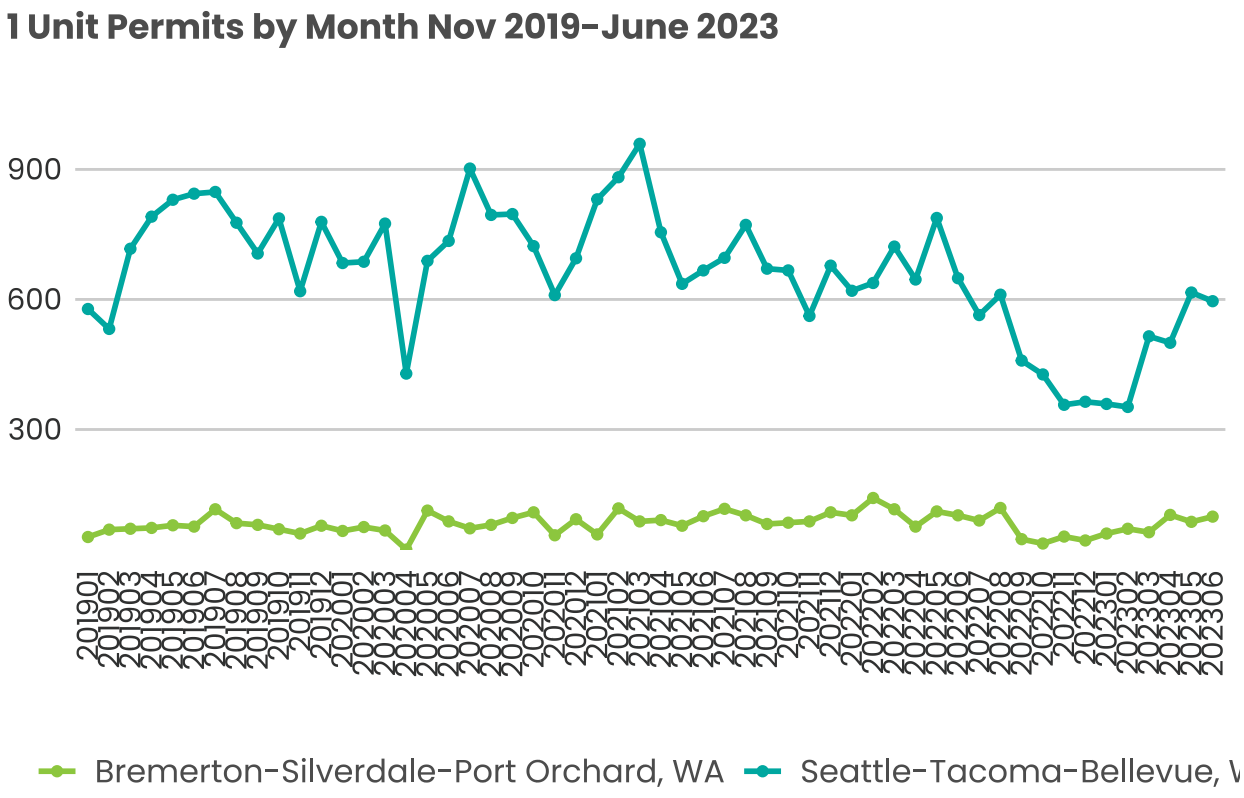
```
df_all<-df_all%>%mutate(MSA_Name=str_trim(Name))
```

```
write.xlsx(df_all, 'census_bps_monthly.xlsx')
```

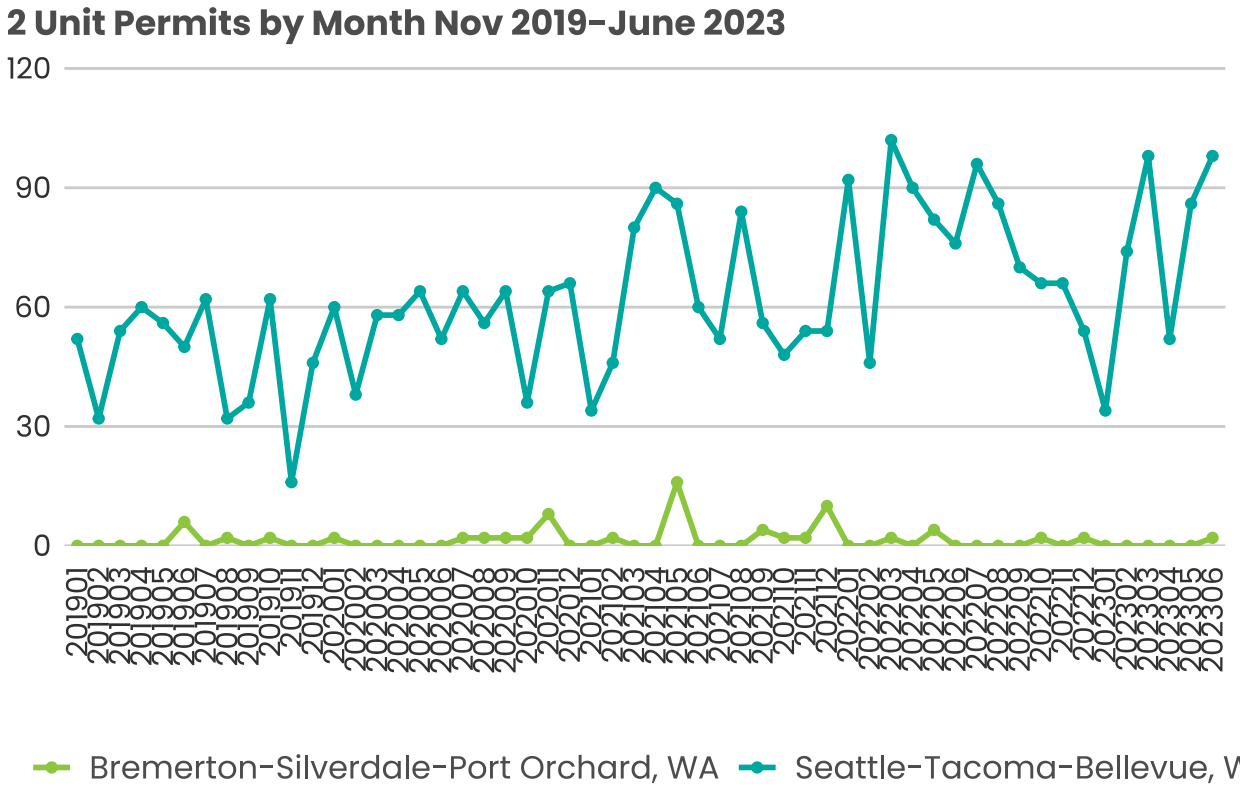
```
static_line_chart(t=df_all, x='Month', y='Total', fill='MSA_Name',title='Total Permits by Month Nov 201
```



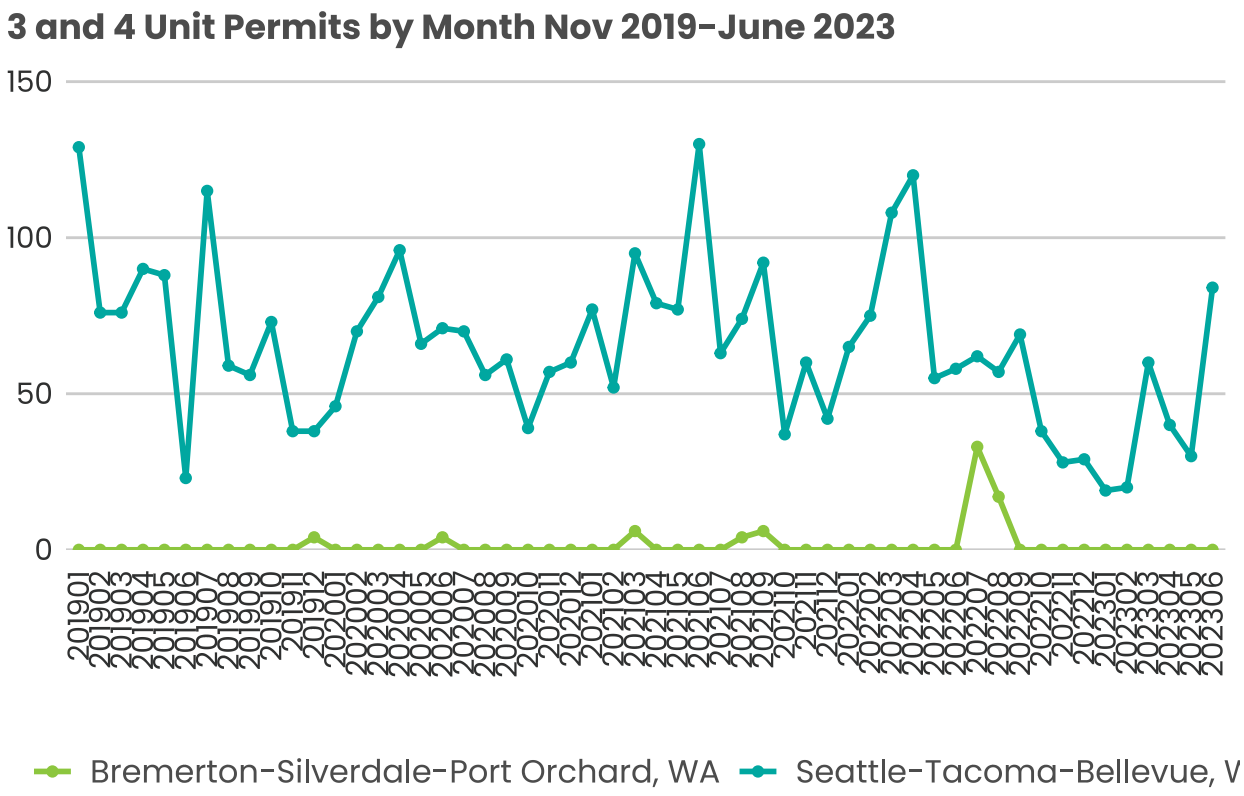
```
static_line_chart(t=df_all, x='Month', y='1 Unit', fill='MSA_Name',  title='1 Unit Permits by Month Nov
```



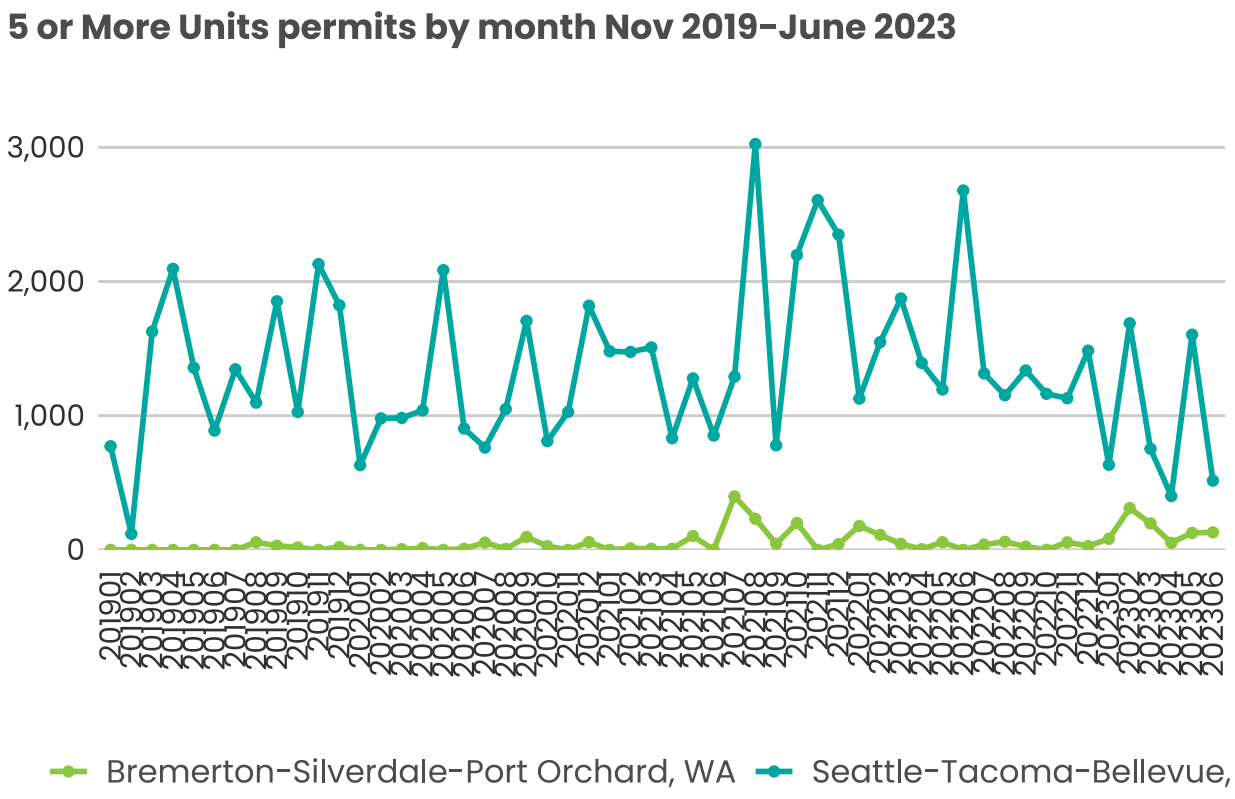

```
static_line_chart(t=df_all, x='Month', y='2 Units', fill='MSA_Name',title='2 Unit Permits by Month Nov
```



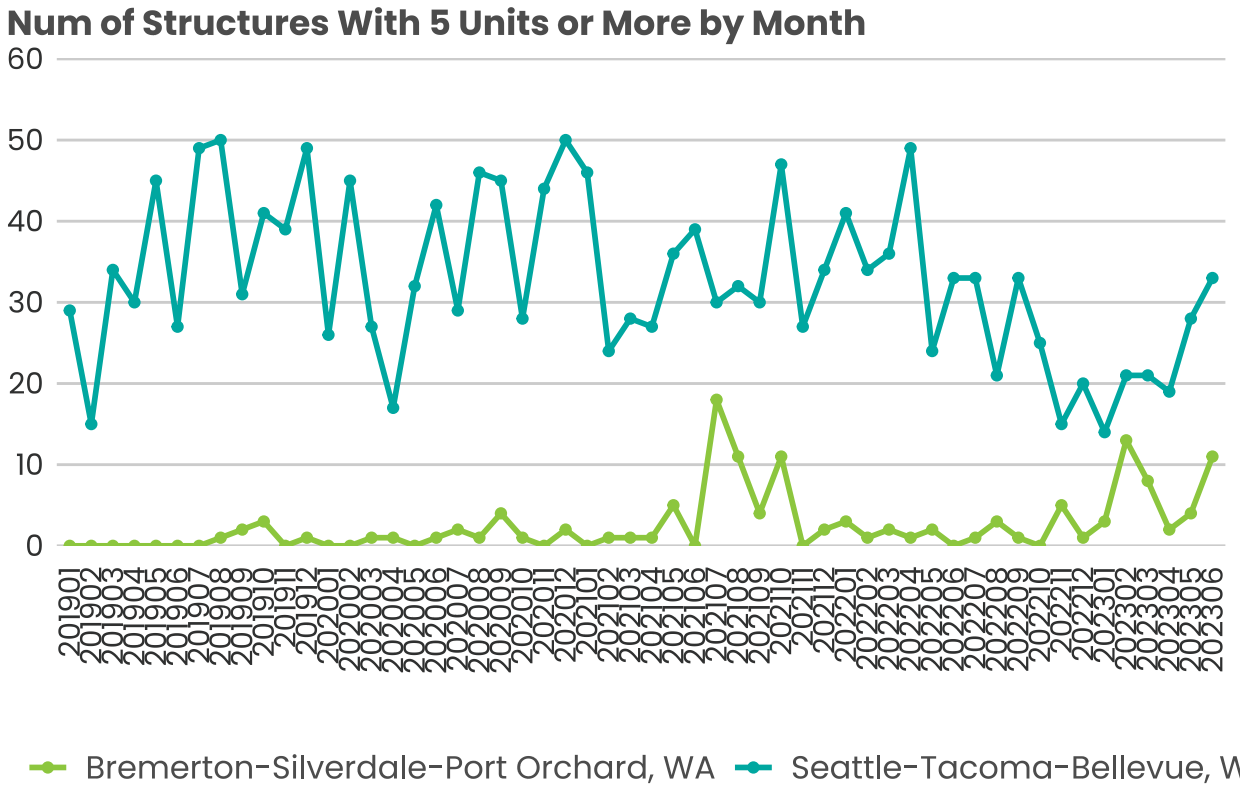

```
static_line_chart(t=df_all, x='Month', y='3 and 4 Units', fill='MSA_Name', title='3 and 4 Unit Permits
```




```
static_line_chart(t=df_all, x='Month', y='5 Units or More', fill='MSA_Name', title='5 or More Units per
```




```
static_line_chart(t=df_all, x='Month', y='Num of Structures With 5 Units or More', title='Num of Structu
```


Num of Structures With 5 Units or More by Month
60
50
40
30
20
10
0
201901
201902
201903
201904
201905
201906
201907
201908
201909
201910
201911
201912
202001
202002
202003
202004
202005
202006
202007
202008
202009
202010
202011
202012
202101
202102
202103
202104
202105
202106
202107
202108
202109
202110
202111
202112
202201
202202
202203
202204
202205
202206
202207
202208
202209
202210
202211
202212
202301
202302
202303
202304
202305
202306
Bremerton-Silverdale-Port Orchard, WA
Seattle-Tacoma-Bellevue, W