In [9]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#https://www.google.com/covid19/mobility/
url='https://drive.google.com/file/d/18gyHbx6rfogq3yQ-GR9COjcGgyYlCnBZ/view?usp=sharing'
url2='https://drive.google.com/uc?id=' + url.split('/')[-2]
df = pd.read_csv(url2, parse_dates=["date"])
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 167657 entries, 0 to 167656
Data columns (total 15 columns):
 #   Column                                         Non-Null Count   Dtype         
---  ------                                         --------------   -----         
 0   country_region_code                            167657 non-null  object        
 1   country_region                                 167657 non-null  object        
 2   sub_region_1                                   167336 non-null  object        
 3   sub_region_2                                   141692 non-null  object        
 4   metro_area                                     0 non-null       float64       
 5   iso_3166_2_code                                25644 non-null   object        
 6   census_fips_code                               0 non-null       float64       
 7   place_id                                       167657 non-null  object        
 8   date                                           167657 non-null  datetime64[ns]
 9   retail_and_recreation_percent_change_from_baseline  101865 non-null  float64       
 10  grocery_and_pharmacy_percent_change_from_baseline  106104 non-null  float64       
 11  parks_percent_change_from_baseline             95186 non-null   float64       
 12  transit_stations_percent_change_from_baseline  87723 non-null   float64       
 13  workplaces_percent_change_from_baseline        158870 non-null  float64       
 14  residential_percent_change_from_baseline       98651 non-null   float64       
dtypes: datetime64[ns](1), float64(8), object(6)
memory usage: 19.2+ MB
```

In [20]:
```
url3='https://drive.google.com/file/d/1Eg8Lffm49bc-bGFkv_4ddrQw8U8WE6P4/view?usp=sharing'
url4='https://drive.google.com/uc?id=' + url3.split('/')[-2]
df1 = pd.read_csv(url4, parse_dates=["date"])
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 158430 entries, 0 to 158429
Data columns (total 15 columns):
 #   Column                                         Non-Null Count   Dtype         
---  ------                                         --------------   -----         
 0   country_region_code                            158430 non-null  object        
 1   country_region                                 158430 non-null  object        
 2   sub_region_1                                   158152 non-null  object        
 3   sub_region_2                                   135654 non-null  object        
 4   metro_area                                     0 non-null       float64       
 5   iso_3166_2_code                                22498 non-null   object        
 6   census_fips_code                               0 non-null       float64       
 7   place_id                                       158430 non-null  object        
 8   date                                           158430 non-null  datetime64[ns]
 9   retail_and_recreation_percent_change_from_baseline  91170 non-null   float64       
 10  grocery_and_pharmacy_percent_change_from_baseline  92489 non-null   float64       
 11  parks_percent_change_from_baseline             87099 non-null   float64       
 12  transit_stations_percent_change_from_baseline  78809 non-null   float64       
 13  workplaces_percent_change_from_baseline        154672 non-null  float64       
 14  residential_percent_change_from_baseline       98407 non-null   float64       
dtypes: datetime64[ns](1), float64(8), object(6)
memory usage: 18.1+ MB
```

SORU1

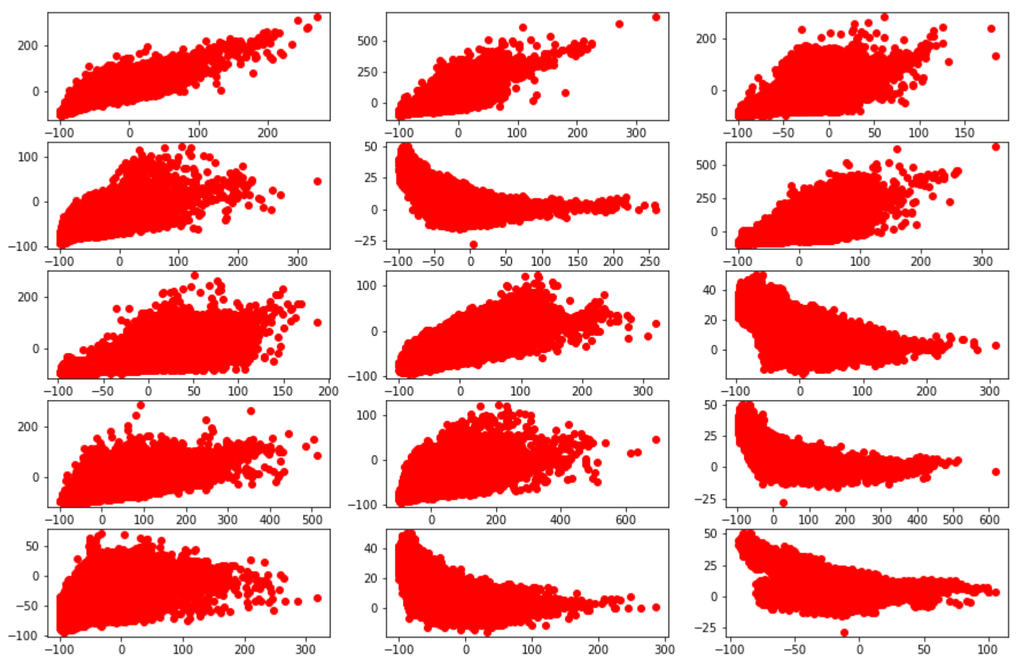
In [11]:
```
df.describe()
```

Out[11]:

	metro_area	census_fips_code	retail_and_recreation_percent_change_from_baseline	grocery_and_pharmacy_percent_change_from_baseline	parks_percent_change_from_baseline	transit_stations_percent_change_from_baseline	workplaces_percent_change_from_baseline	residential_percent_change_from_baseline
count	0.0	0.0	101865.000000	106104.000000	95186.000000	87723.000000	158870.000000	98651.000000
mean	NaN	NaN	-26.758749	2.700699	5.780503	-25.233006	-19.002795	6.990188
std	NaN	NaN	30.125282	32.575277	53.051874	35.328635	21.563078	8.715291
min	NaN	NaN	-100.000000	-100.000000	-100.000000	-100.000000	-94.000000	-28.000000
25%	NaN	NaN	-44.000000	-9.000000	-26.000000	-48.000000	-30.000000	1.000000
50%	NaN	NaN	-24.000000	5.000000	2.000000	-25.000000	-17.000000	5.000000
75%	NaN	NaN	-8.000000	18.000000	30.000000	-5.000000	-6.000000	12.000000
max	NaN	NaN	333.000000	321.000000	694.000000	318.000000	136.000000	50.000000

SORU2

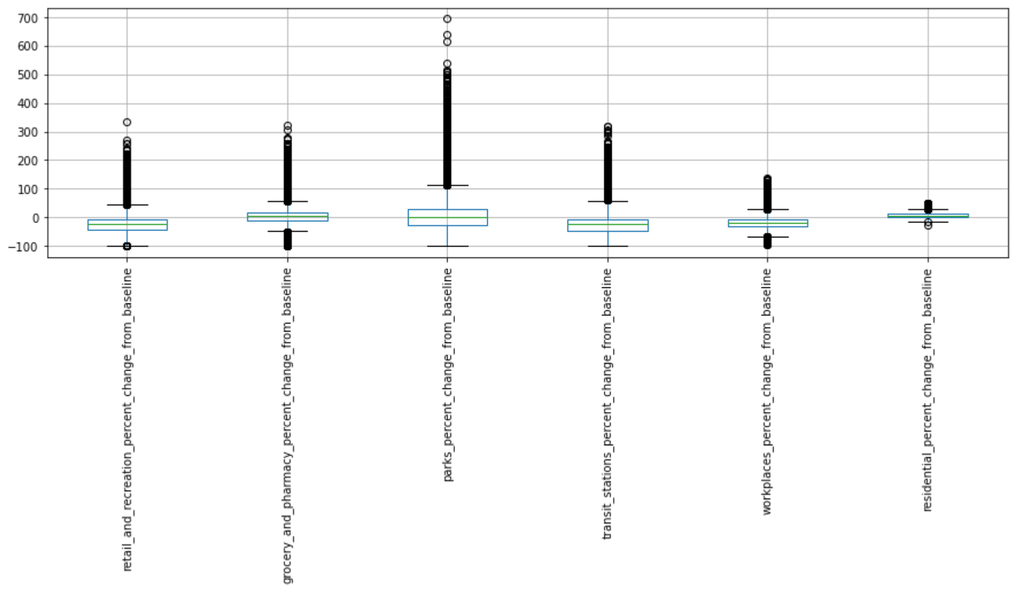
In [12]:
```
import matplotlib.pyplot as plt
fig, axes = plt.subplots(5, 3, figsize=(15,10))
index = 0
for i in range(9, 14):
    for j in range(i+1, 15):
        ax1 = int(index/3)
        ax2 = index % 3
        axes[ax1][ax2].scatter(df[df.columns[i]], df[df.columns[j]], color='red')
        #axes[ax1][ax2].set_xlabel(df.columns[i])
        #axes[ax1][ax2].set_ylabel(df.columns[j])
        index = index + 1
```



[0,0], [0,1], [0,3], [1,0], [1,0], [1,2], [2,0], [2,1], [3,0], [3,1] pozitif yönlü korelasyon vardır. [1,1], [2,2], [3,2], [4,1], [4,2] negatif yönlü korelasyon vardır.

SORU3

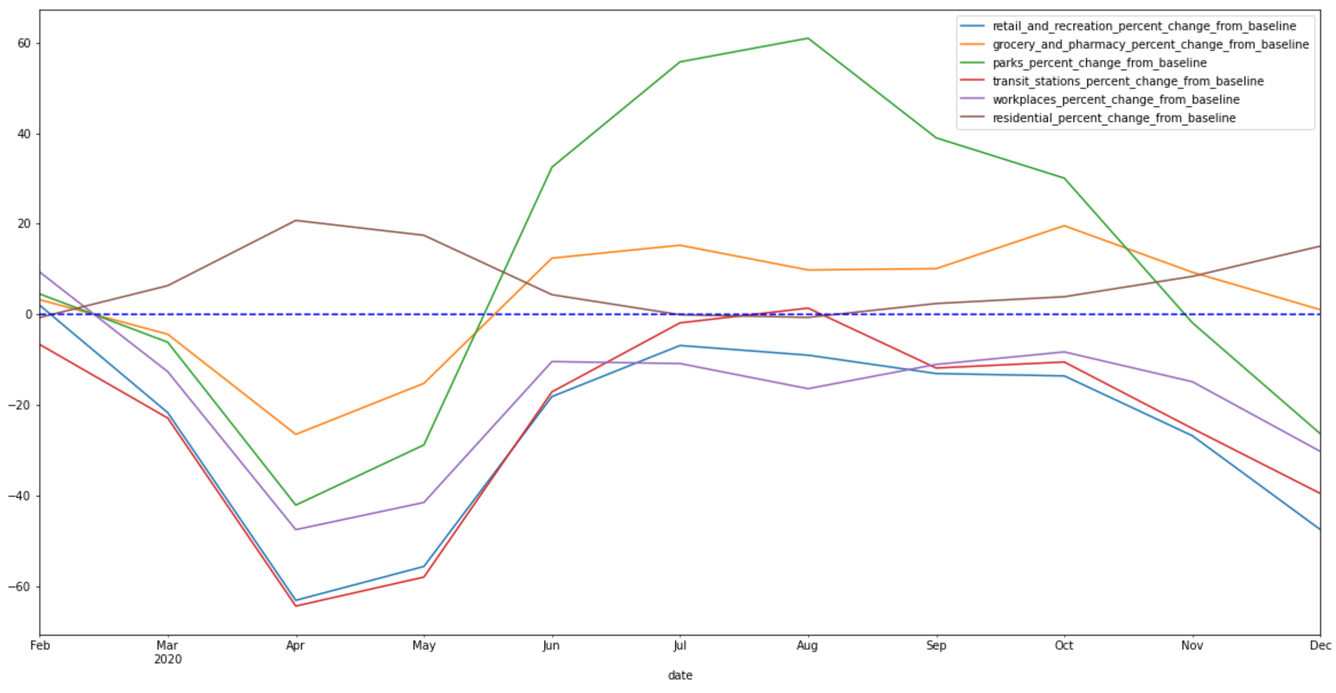
In [15]:
```
import numpy as np
np.random.seed(1234)
plt.figure(figsize=(15,4))
boxplot = df.boxplot(column=["retail_and_recreation_percent_change_from_baseline",
                             "grocery_and_pharmacy_percent_change_from_baseline",
                             "parks_percent_change_from_baseline",
                             "transit_stations_percent_change_from_baseline",
                             "workplaces_percent_change_from_baseline",
                             "residential_percent_change_from_baseline"])
plt.xticks(rotation=90)
plt.show()
```



Bütün boxplotlarda outlier vardır.

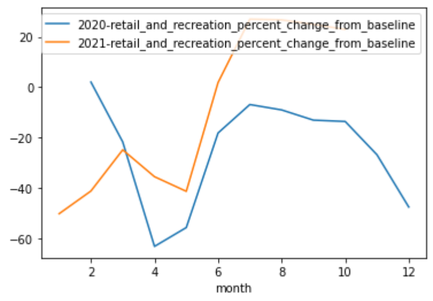
SORU4

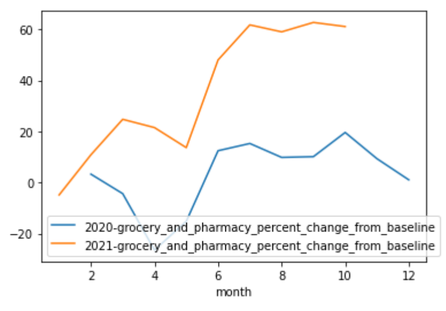
In [16]:
```
import numpy as np
df.index = pd.to_datetime(df[df.columns[8]])
monthly = df.groupby(pd.Grouper(freq='M')).mean()
del monthly["metro_area"]
del monthly["census_fips_code"]
ax = monthly.plot(kind='line', figsize=(20,10))
df['baseline'] = 0
df['baseline'].plot(linestyle="--", color="blue")
plt.show()
```

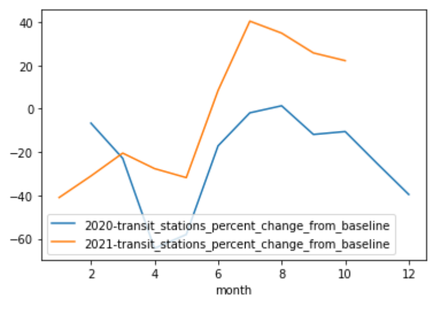


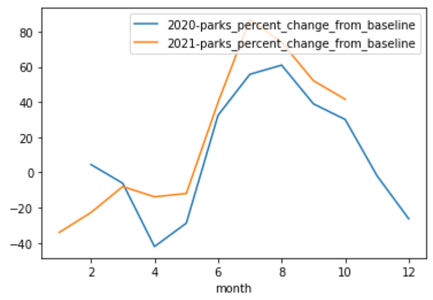
SORU5

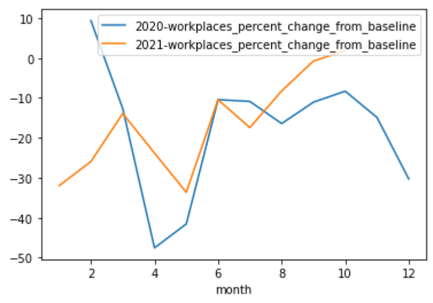
In [18]:
```
import datetime

df['month'] = pd.DatetimeIndex(df['date']).month
df1['month'] = pd.DatetimeIndex(df1['date']).month

df.index = pd.to_datetime(df[df.columns[8]])
df1.index = pd.to_datetime(df1[df1.columns[8]])

monthly1 = df.groupby(pd.Grouper(freq='M')).mean()
monthly2 = df1.groupby(pd.Grouper(freq='M')).mean()

del monthly1['metro_area']
del monthly1['census_fips_code']
del monthly2['metro_area']
del monthly2['census_fips_code']

ax = monthly1.plot(y="retail_and_recreation_percent_change_from_baseline", x='month', kind='line',label="2020-retail_and_recreation_percent_change_from_baseline")
monthly2.plot(ax=ax, y="retail_and_recreation_percent_change_from_baseline", x='month', kind='line',label="2021-retail_and_recreation_percent_change_from_baseline")

ax = monthly1.plot(y="grocery_and_pharmacy_percent_change_from_baseline", x='month', kind='line',label="2020-grocery_and_pharmacy_percent_change_from_baseline")
monthly2.plot(ax=ax, y="grocery_and_pharmacy_percent_change_from_baseline", x='month', kind='line',label="2021-grocery_and_pharmacy_percent_change_from_baseline")

ax = monthly1.plot(y="transit_stations_percent_change_from_baseline", x='month', kind='line',label="2020-transit_stations_percent_change_from_baseline")
monthly2.plot(ax=ax, y="transit_stations_percent_change_from_baseline", x='month', kind='line',label="2021-transit_stations_percent_change_from_baseline")

ax = monthly1.plot(y="parks_percent_change_from_baseline", x='month', kind='line',label="2020-parks_percent_change_from_baseline")
monthly2.plot(ax=ax, y="parks_percent_change_from_baseline", x='month', kind='line',label="2021-parks_percent_change_from_baseline")

ax = monthly1.plot(y="workplaces_percent_change_from_baseline", x='month', kind='line',label="2020-workplaces_percent_change_from_baseline")
monthly2.plot(ax=ax, y="workplaces_percent_change_from_baseline", x='month', kind='line',label="2021-workplaces_percent_change_from_baseline")

ax = monthly1.plot(y="residential_percent_change_from_baseline", x='month', kind='line',label="2020-residential_percent_change_from_baseline")
monthly2.plot(ax=ax, y="residential_percent_change_from_baseline", x='month', kind='line',label="2021-residential_percent_change_from_baseline")

plt.show()
```

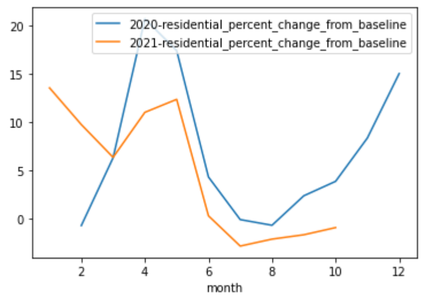












KAYNAKÇA:
https://www.educative.io/edpresso/how-to-delete-a-column-in-pandas
http://www.cse.msu.edu/~ptan/dmbook/tutorials/tutorial3/tutorial3.html
https://pandas.pydata.org/pandas-docs/dev/reference/api/pandas.plotting.boxplot.html
https://stackoverflow.com/questions/10998621/rotate-axis-text-in-python-matplotlib
http://www.cse.msu.edu/~ptan/dmbook/tutorials/tutorial4/tutorial4.html
https://stackoverflow.com/questions/25146121/extracting-just-month-and-year-separately-from-pandas-datetime-column?rq=1