

```
In [48]: # Imports
import pandas as pd
import numpy as np

import matplotlib
import matplotlib.pyplot as plt

import seaborn as sns

from scipy.stats import sem
```

```
In [49]: # Raw data import
ct_df = pd.read_csv("data/clinical_trial_data.csv")
md_df = pd.read_csv("data/mouse_drug_data.csv")

merged_df = pd.merge(ct_df, md_df, on="Mouse ID")
merged_df.head()
```

Out[49]:

| | Mouse ID | Timepoint | Tumor Volume (mm3) | Metastatic Sites | Drug |
|---|----------|-----------|--------------------|------------------|-----------|
| 0 | b128 | 0 | 45.000000 | 0 | Capomulin |
| 1 | b128 | 5 | 45.651331 | 0 | Capomulin |
| 2 | b128 | 10 | 43.270852 | 0 | Capomulin |
| 3 | b128 | 15 | 43.784893 | 0 | Capomulin |
| 4 | b128 | 20 | 42.731552 | 0 | Capomulin |

```
In [50]: # Data cleaning and engineering/grouping pre-process before visualization

merged_df.loc[(merged_df['Drug'] != 'Ceftamin') & \
              (merged_df['Drug'] != 'Naftisol') & \
              (merged_df['Drug'] != 'Naftisol') & \
              (merged_df['Drug'] != 'Propriova') & \
              (merged_df['Drug'] != 'Ramicane') & \
              (merged_df['Drug'] != 'Stelasyn') & \
              (merged_df['Drug'] != 'Zoniferol') ,:]

group_drug_df = merged_df.groupby(['Drug', 'Timepoint'])
TumorVolumeMean_df = group_drug_df.mean()['Tumor Volume (mm3)']
TumorVolumeMean_df.head()

TVM_indexreset_df = TumorVolumeMean_df.reset_index()
TVM_indexreset_df.head()

CapomulinMeanVolume = TVM_indexreset_df.loc[TVM_indexreset_df['Drug'] == 'Capo
mulin']
InfubinolMeanVolume = TVM_indexreset_df.loc[TVM_indexreset_df['Drug'] == 'Infu
binol']
KetaprilMeanVolume = TVM_indexreset_df.loc[TVM_indexreset_df['Drug'] == 'Ketap
ril']
PlaceboMeanVolume = TVM_indexreset_df.loc[TVM_indexreset_df['Drug'] == 'Placeb
o']
```

Tumor volume change over time per treatment

```

In [51]: fig, ax = plt.subplots()

plt.scatter(CapomulinMeanVolume['Timepoint'],CapomulinMeanVolume['Tumor Volume (mm3)'], label = 'Capomulin', marker="o", facecolors="red", edgecolors="black", alpha=0.75)
plt.scatter(InfubinolMeanVolume['Timepoint'],InfubinolMeanVolume['Tumor Volume (mm3)'], label = 'Infubinol',marker="^", facecolors="blue", edgecolors="black", alpha=0.75)
plt.scatter(KetaprilMeanVolume['Timepoint'],KetaprilMeanVolume['Tumor Volume (mm3)'], label = 'Ketapril',marker="s", facecolors="green", edgecolors="black", alpha=0.75)
plt.scatter(PlaceboMeanVolume['Timepoint'],PlaceboMeanVolume['Tumor Volume (mm3)'], label = 'Placebo',marker="D", facecolors="black", edgecolors="black", alpha=0.75)

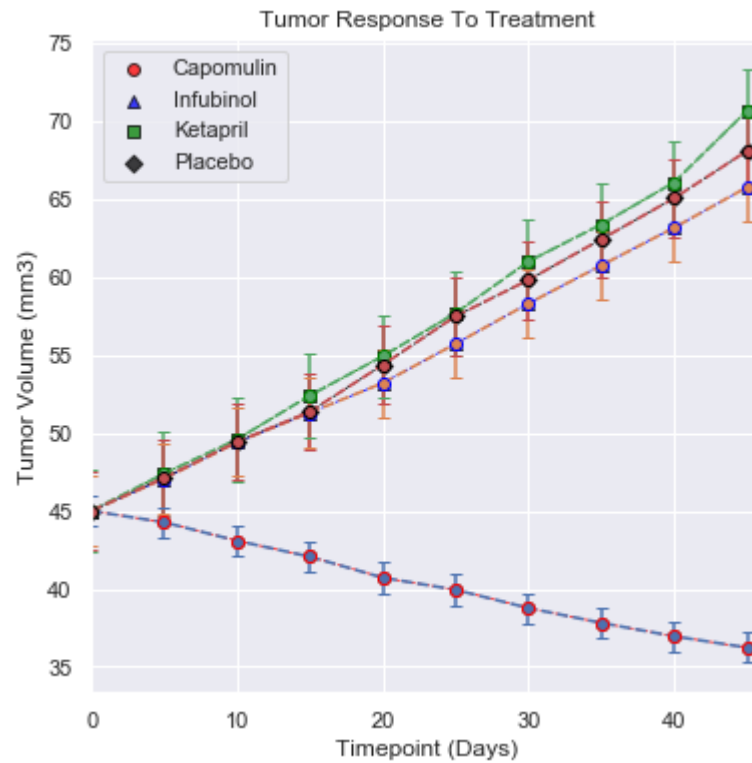
plt.xlim(0, max(PlaceboMeanVolume['Timepoint']+1))
plt.title('Tumor Response To Treatment')
plt.xlabel("Timepoint (Days)")
plt.ylabel("Tumor Volume (mm3)")
plt.legend(loc="best")
plt.grid(b=True,axis='both')

plt.plot(CapomulinMeanVolume['Timepoint'],CapomulinMeanVolume['Tumor Volume (mm3)'], linestyle='--',linewidth=0.7, color="red")
plt.plot(InfubinolMeanVolume['Timepoint'],InfubinolMeanVolume['Tumor Volume (mm3)'], linestyle='--',linewidth=0.7,color="blue")
plt.plot(KetaprilMeanVolume['Timepoint'],KetaprilMeanVolume['Tumor Volume (mm3)'], linestyle='--',linewidth=0.7,color="green")
plt.plot(PlaceboMeanVolume['Timepoint'],PlaceboMeanVolume['Tumor Volume (mm3)'], linestyle='--',linewidth=0.7,color="black")

plt.errorbar(CapomulinMeanVolume['Timepoint'], CapomulinMeanVolume['Tumor Volume (mm3)'], yerr = sem(CapomulinMeanVolume['Tumor Volume (mm3)']),linestyle = "--",fmt = 'o',capsize=3,capthick=1,markeredgecolor='red')
plt.errorbar(InfubinolMeanVolume['Timepoint'],InfubinolMeanVolume['Tumor Volume (mm3)'], yerr = sem(InfubinolMeanVolume['Tumor Volume (mm3)']),linestyle = "--",fmt = 'o',capsize=3,capthick=1,markeredgecolor='blue')
plt.errorbar(KetaprilMeanVolume['Timepoint'],KetaprilMeanVolume['Tumor Volume (mm3)'], yerr = sem(KetaprilMeanVolume['Tumor Volume (mm3)']),linestyle = "--",fmt = 'o',capsize=3,capthick=1,markeredgecolor='green')
plt.errorbar(PlaceboMeanVolume['Timepoint'],PlaceboMeanVolume['Tumor Volume (mm3)'], yerr = sem(PlaceboMeanVolume['Tumor Volume (mm3)']),linestyle = "--",fmt = 'o',capsize=3,capthick=1,markeredgecolor='black')

plt.savefig('reports/figures/tumor_response.png')
plt.show()

```



Metastatic site change over time

```

In [52]: fig, ax = plt.subplots()
metaStaticMean_df = group_drug_df.mean()['Metastatic Sites']
MSS_indexreset_df = metaStaticMean_df.reset_index()
MSS_indexreset_df.head()

CapomulinMeanVolumeMSS = MSS_indexreset_df.loc[MSS_indexreset_df['Drug'] == 'C
apomulin']
InfubinolMeanVolumeMSS = MSS_indexreset_df.loc[MSS_indexreset_df['Drug'] == 'I
nfubinol']
KetaprilMeanVolumeMSS = MSS_indexreset_df.loc[MSS_indexreset_df['Drug'] == 'Ke
tapril']
PlaceboMeanVolumeMSS = MSS_indexreset_df.loc[MSS_indexreset_df['Drug'] == 'Pla
cebo']

plt.scatter(CapomulinMeanVolumeMSS['Timepoint'],CapomulinMeanVolumeMSS['Metast
atic Sites'], label = 'Capomulin', marker="o", facecolors="red", edgecolors="b
lack",alpha=0.75)
plt.scatter(InfubinolMeanVolumeMSS['Timepoint'],InfubinolMeanVolumeMSS['Metast
atic Sites'], label = 'Infubinol',marker="^", facecolors="blue", edgecolors="b
lack",alpha=0.75)
plt.scatter(KetaprilMeanVolumeMSS['Timepoint'],KetaprilMeanVolumeMSS['Metastat
ic Sites'], label = 'Ketapril',marker="s", facecolors="green", edgecolors="bla
ck",alpha=0.75)
plt.scatter(PlaceboMeanVolumeMSS['Timepoint'],PlaceboMeanVolumeMSS['Metastatic
Sites'], label = 'Placebo',marker="D", facecolors="black", edgecolors="black",
alpha=0.75)

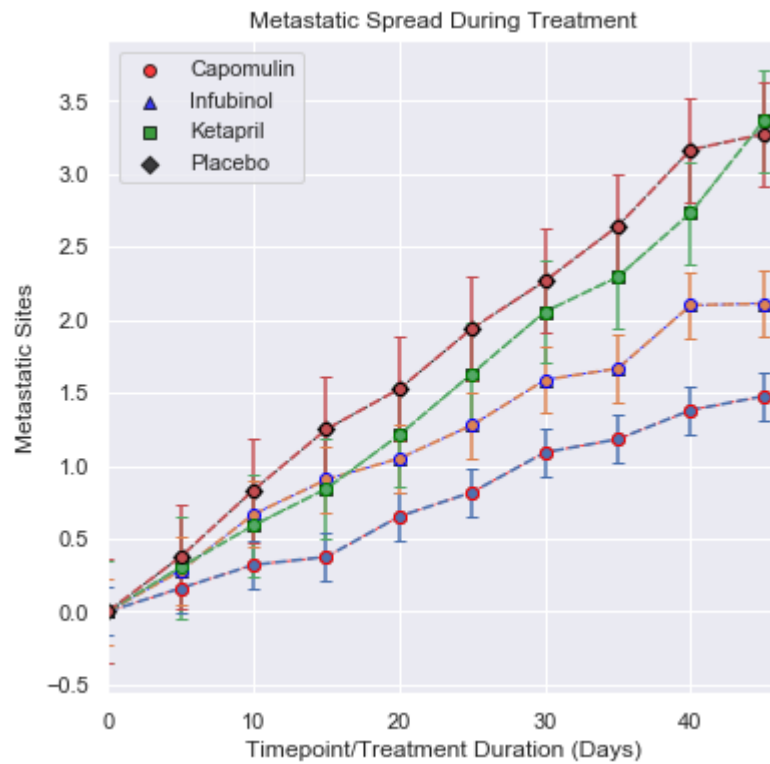
plt.xlim(0, max(PlaceboMeanVolumeMSS['Timepoint']+1))
plt.title('Metastatic Spread During Treatment')
plt.xlabel("Timepoint/Treatment Duration (Days)")
plt.ylabel("Metastatic Sites")
plt.legend(loc="best")
plt.grid(b=True,axis='both')

plt.plot(CapomulinMeanVolumeMSS['Timepoint'],CapomulinMeanVolumeMSS['Metastati
c Sites'], linestyle='--',linewidth=0.7, color="red")
plt.plot(InfubinolMeanVolumeMSS['Timepoint'],InfubinolMeanVolumeMSS['Metastati
c Sites'], linestyle='--',linewidth=0.7,color="blue")
plt.plot(KetaprilMeanVolumeMSS['Timepoint'],KetaprilMeanVolumeMSS['Metastatic
Sites'], linestyle='--',linewidth=0.7,color="green")
plt.plot(PlaceboMeanVolumeMSS['Timepoint'],PlaceboMeanVolumeMSS['Metastatic Si
tes'], linestyle='--',linewidth=0.7,color="black")

plt.errorbar(CapomulinMeanVolumeMSS['Timepoint'], CapomulinMeanVolumeMSS['Meta
static Sites'], yerr = sem(CapomulinMeanVolumeMSS['Metastatic Sites']),linesty
le = "--",fmt = 'o',capsize=3,capthick=1,markeredgcolor='red')
plt.errorbar(InfubinolMeanVolumeMSS['Timepoint'],InfubinolMeanVolumeMSS['Metas
tatic Sites'], yerr = sem(InfubinolMeanVolumeMSS['Metastatic Sites']),linestyl
e = "--",fmt = 'o',capsize=3,capthick=1,markeredgcolor='blue')
plt.errorbar(KetaprilMeanVolumeMSS['Timepoint'],KetaprilMeanVolumeMSS['Metasta
tic Sites'], yerr = sem(KetaprilMeanVolumeMSS['Metastatic Sites']),linestyle
="--",fmt = 'o',capsize=3,capthick=1,markeredgcolor='green')
plt.errorbar(PlaceboMeanVolumeMSS['Timepoint'],PlaceboMeanVolumeMSS['Metastati
c Sites'], yerr = sem(PlaceboMeanVolumeMSS['Metastatic Sites']),linestyle ="-
-",fmt = 'o',capsize=3,capthick=1,markeredgcolor='black')

```

```
plt.savefig('reports/figures/metastatic_spread.png')  
plt.show()
```



Survival rate over time

```

In [53]: fig, ax = plt.subplots()

SR_df = merged_df.groupby(['Drug', 'Timepoint']).count()['Mouse ID']
SR_indexreset_df = SR_df.reset_index()
SR_indexreset_df.head()

CapomulinMeanVolumeSR = SR_indexreset_df.loc[SR_indexreset_df['Drug'] == 'Capo
mulin']
InfubinolMeanVolumeSR = SR_indexreset_df.loc[SR_indexreset_df['Drug'] == 'Infu
binol']
KetaprilMeanVolumeSR = SR_indexreset_df.loc[SR_indexreset_df['Drug'] == 'Ketap
ril']
PlaceboMeanVolumeSR = SR_indexreset_df.loc[SR_indexreset_df['Drug'] == 'Placeb
o']

def testfunc(num1):
    num1 = float(num1)
    percentage = num1/26
    return percentage

SR_indexreset_df= pd.pivot_table(SR_indexreset_df, index='Timepoint', columns=
'Drug', values='Mouse ID', aggfunc = testfunc)
SR_indexreset_df= SR_indexreset_df

plt.scatter(CapomulinMeanVolumeSR['Timepoint'],CapomulinMeanVolumeSR['Mouse I
D'], label = 'Capomulin', marker="o", facecolors="red", edgecolors="black",alp
ha=0.75)
plt.scatter(InfubinolMeanVolumeSR['Timepoint'],InfubinolMeanVolumeSR['Mouse I
D'], label = 'Infubinol',marker="^", facecolors="blue", edgecolors="black",alp
ha=0.75)
plt.scatter(KetaprilMeanVolumeSR['Timepoint'],KetaprilMeanVolumeSR['Mouse ID'
], label = 'Ketapril',marker="s", facecolors="green", edgecolors="black",alpha
=0.75)
plt.scatter(PlaceboMeanVolumeSR['Timepoint'],PlaceboMeanVolumeSR['Mouse ID'],
label = 'Placebo',marker="D", facecolors="black", edgecolors="black",alpha=0.7
5)

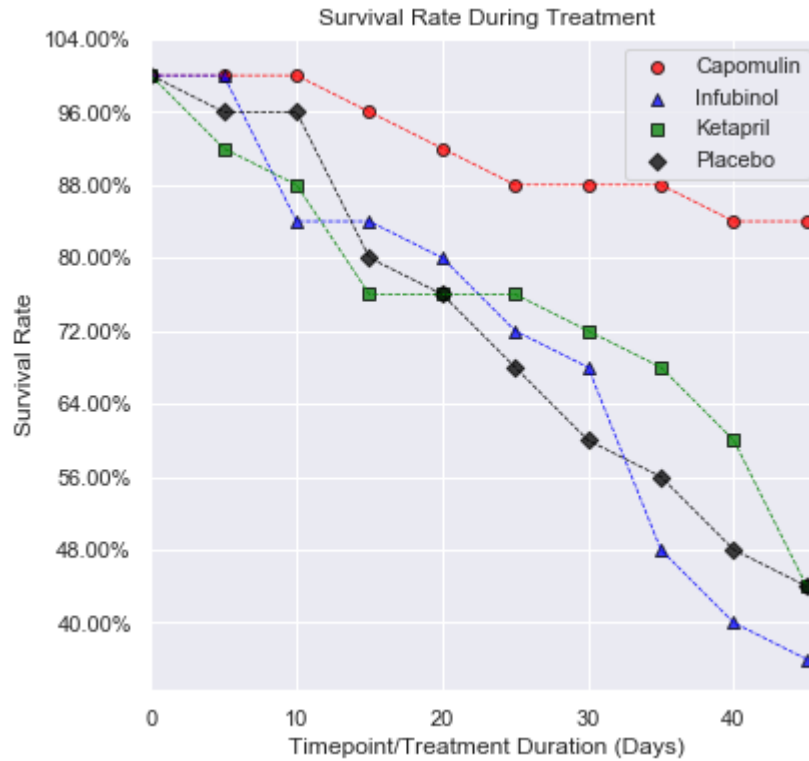
plt.xlim(0, max(PlaceboMeanVolumeSR['Timepoint'])+1))
plt.title('Survival Rate During Treatment')
plt.xlabel('Timepoint/Treatment Duration (Days)')
plt.ylabel("Survival Rate")
plt.legend(loc="best")

plt.plot(CapomulinMeanVolumeSR['Timepoint'],CapomulinMeanVolumeSR['Mouse ID'],
linestyle='--',linewidth=0.7, color="red")
plt.plot(InfubinolMeanVolumeSR['Timepoint'],InfubinolMeanVolumeSR['Mouse ID'],
linestyle='--',linewidth=0.7,color="blue")
plt.plot(KetaprilMeanVolumeSR['Timepoint'],KetaprilMeanVolumeSR['Mouse ID'], l
inestyle='--',linewidth=0.7,color="green")
plt.plot(PlaceboMeanVolumeSR['Timepoint'],PlaceboMeanVolumeSR['Mouse ID'], lin
estyle='--',linewidth=0.7,color="black")

vals = ax.get_yticks()
ax.set_yticklabels(['{:3.2f}%'.format(x*4) for x in vals])
SR_indexreset_df.head()

```

```
plt.savefig('reports/figures/survival_rate.png')
plt.show()
```



In []:

```
In [54]: #Calculate percentage change for tumor volume chart
def return_pc_45(df):
    cmvc0 =df.loc[df['Timepoint'] == 0,'Tumor Volume (mm3)']
    cmvc45=df.loc[df['Timepoint'] == 45,'Tumor Volume (mm3)']
    cmvc_percentchange=(cmvc0.values[0] - cmvc45.values[0])/cmvc0.values[0]*100
    return np.round(cmvc_percentchange,decimals=2)

print(
return_pc_45(CapomulinMeanVolume),
return_pc_45(PlaceboMeanVolume),
return_pc_45(InfubinolMeanVolume),
return_pc_45(KetaprilMeanVolume)
)
```

19.48 -51.3 -46.12 -57.03


```
In [55]: pc_45_list = [return_pc_45(CapomulinMeanVolume),return_pc_45(PlaceboMeanVolume),\
                    return_pc_45(InfubinolMeanVolume),return_pc_45(KetaprilMeanVolume)]
print(pc_45_list)

#Switch negative and positive for chart
pc_45_list=np.negative(pc_45_list)
print(pc_45_list)

#Color list based upon value
colors = []
for value in pc_45_list:
    if value < 0:
        colors.append('red')
    else:
        colors.append('green')

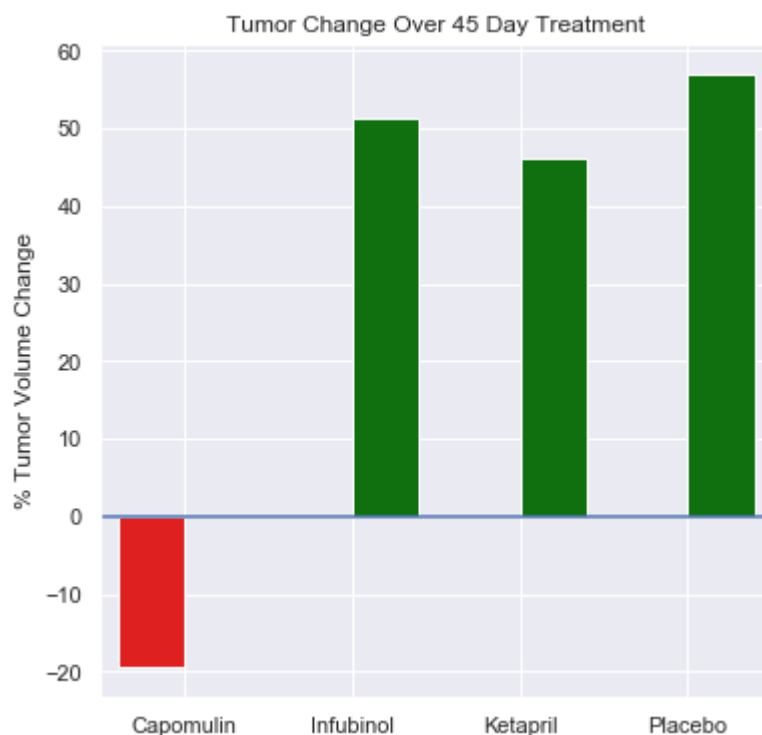
print(colors)

[19.48, -51.3, -46.12, -57.03]
[-19.48  51.3  46.12  57.03]
['red', 'green', 'green', 'green']
```

Tumor Change Over 45 Day Treatment

```
In [56]: # Bar graph comparing total % tumor volume change for each drug across the full 45 days
x=['Capomulin','Infubinol','Ketapril','Placebo']
y=pc_45_list
fig, ax = plt.subplots()
sns.set(rc={'figure.figsize':(6,6)})
sns.barplot(x,y,order=x, palette=colors, hue = colors)
ax.set_title("Tumor Change Over 45 Day Treatment")
ax.legend_.remove()
plt.grid(b=True,axis='both')
plt.axhline(y=0, color='b', linestyle='--')
plt.ylabel("% Tumor Volume Change")

plt.savefig('reports/figures/tumor_change.png')
plt.show()
```



In []:

```

In [57]: #Debugging
#https://pandas.pydata.org/pandas-docs/stable/reshaping.html
#https://pandas.pydata.org/pandas-docs/stable/generated/pandas.pivot_table.html
def testfunc(num1):
    num1 = float(num1)
    percentage = num1/26
    return percentage

SR_df = merged_df.groupby(['Drug', 'Timepoint']).count()['Mouse ID']
SR_indexreset_df = SR_df.reset_index()

SR_indexreset_df= pd.pivot_table(SR_indexreset_df, index='Timepoint', columns=
'Drug', values='Mouse ID', aggfunc = testfunc)

SR_indexreset_df= SR_indexreset_df
SR_indexreset_df.head(45)

```

Out[57]:

| | Drug | Capomulin | Ceftamin | Infubinol | Ketapril | Naftisol | Placebo | Propriva | Ramicane | Σ |
|-----------|------|-----------|----------|-----------|----------|----------|----------|----------|----------|---|
| Timepoint | | | | | | | | | | |
| 0 | | 0.961538 | 0.961538 | 0.961538 | 0.961538 | 0.961538 | 0.961538 | 1.000000 | 0.961538 | 1 |
| 5 | | 0.961538 | 0.807692 | 0.961538 | 0.884615 | 0.884615 | 0.923077 | 0.961538 | 0.961538 | 0 |
| 10 | | 0.961538 | 0.769231 | 0.807692 | 0.846154 | 0.807692 | 0.923077 | 0.884615 | 0.923077 | 0 |
| 15 | | 0.923077 | 0.730769 | 0.807692 | 0.730769 | 0.807692 | 0.769231 | 0.653846 | 0.923077 | 0 |
| 20 | | 0.884615 | 0.692308 | 0.769231 | 0.730769 | 0.769231 | 0.730769 | 0.653846 | 0.884615 | 0 |
| 25 | | 0.846154 | 0.692308 | 0.692308 | 0.730769 | 0.692308 | 0.653846 | 0.538462 | 0.884615 | 0 |
| 30 | | 0.846154 | 0.615385 | 0.653846 | 0.692308 | 0.576923 | 0.576923 | 0.500000 | 0.884615 | 0 |
| 35 | | 0.846154 | 0.538462 | 0.461538 | 0.653846 | 0.576923 | 0.538462 | 0.384615 | 0.807692 | 0 |
| 40 | | 0.807692 | 0.538462 | 0.384615 | 0.576923 | 0.576923 | 0.461538 | 0.346154 | 0.769231 | 0 |
| 45 | | 0.807692 | 0.500000 | 0.346154 | 0.423077 | 0.500000 | 0.423077 | 0.269231 | 0.769231 | 0 |

```
In [58]: # #https://matplotlib.org/1.2.1/examples/pylab_examples/errorbar_demo.html
# plt.scatter(CapomulinMeanVolume['Timepoint'],CapomulinMeanVolume['Tumor Volume (mm3)'], label = 'Capomulin', marker="o", facecolors="red", edgecolors="black",alpha=0.75)
# plt.scatter(InfubinolMeanVolume['Timepoint'],InfubinolMeanVolume['Tumor Volume (mm3)'], label = 'Infubinol',marker="^", facecolors="blue", edgecolors="black",alpha=0.75)
# plt.scatter(KetaprilMeanVolume['Timepoint'],KetaprilMeanVolume['Tumor Volume (mm3)'], label = 'Ketapril',marker="s", facecolors="green", edgecolors="black",alpha=0.75)
# plt.scatter(PlaceboMeanVolume['Timepoint'],PlaceboMeanVolume['Tumor Volume (mm3)'], label = 'Placebo',marker="D", facecolors="black", edgecolors="black",alpha=0.75)

# group_drug_df = merged_df.groupby(['Drug','Timepoint'])
# TumorVolumeMean_df = group_drug_df.mean()['Tumor Volume (mm3)']
# TumorVolumeMean_df.head()

# TVM_indexreset_df = TumorVolumeMean_df.reset_index()
# TVM_indexreset_df.head()

# CapomulinMeanVolume = TVM_indexreset_df.loc[TVM_indexreset_df['Drug'] == 'Capomulin']
# InfubinolMeanVolume = TVM_indexreset_df.loc[TVM_indexreset_df['Drug'] == 'Infubinol']
# KetaprilMeanVolume = TVM_indexreset_df.loc[TVM_indexreset_df['Drug'] == 'Ketapril']
# PlaceboMeanVolume = TVM_indexreset_df.loc[TVM_indexreset_df['Drug'] == 'Placebo']

# plt.grid(b=True,axis='both')
# plt.show()
```

In []: