DSC 640: Weeks 5 – 6
Author: Kimberly Cable
Date: Oct 8, 2022
Exercise 3.2 Charts

Treemap

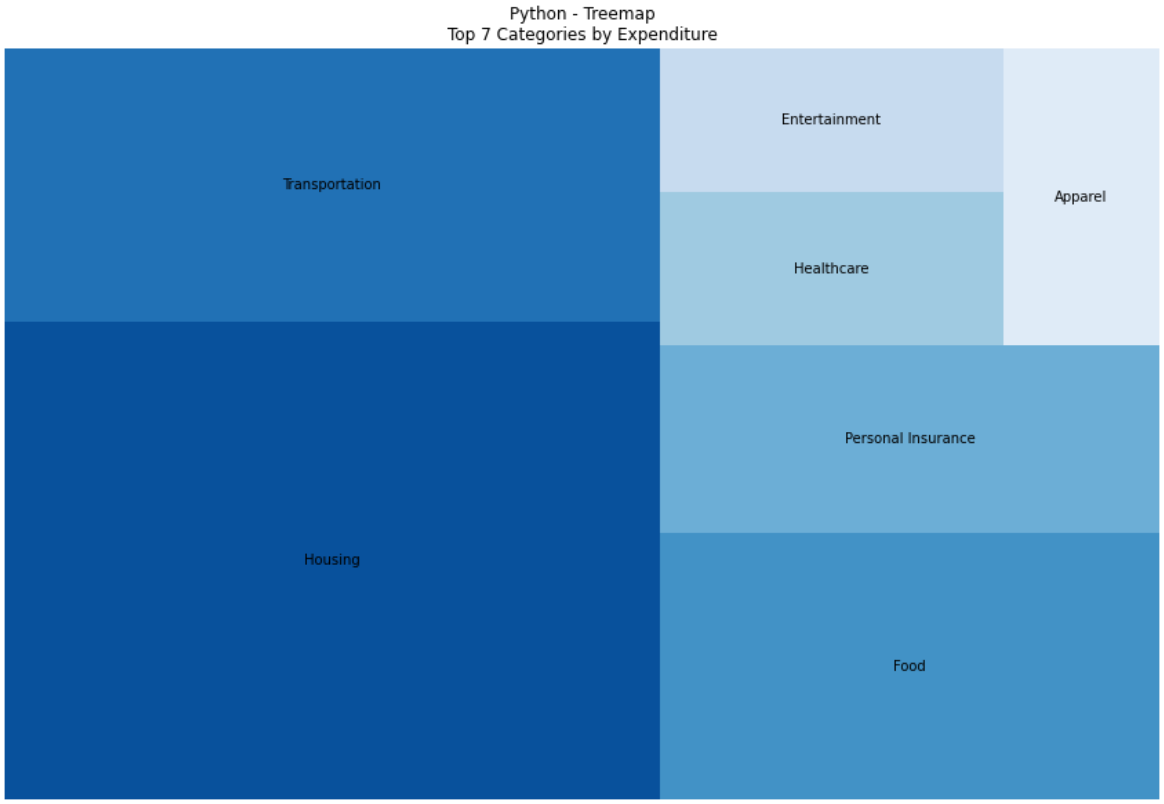
Python

```
# get top 7 categories by expenditures
category_year = expenditures_df.groupby('category')['expenditure'].sum().nlargest(7)
category_year
```

```
fig, ax = plt.subplots(figsize = (15, 10))
squarify.plot(sizes = category_year.values, label = category_year.index,
          color = sb.color_palette("Blues_r", len(category_year.values)))
plt.title('Python - Treemap\nTop 7 Categories by Expenditure')
ax.axis('off')

plt.show()

# Save figure
ax.get_figure().savefig('images/treemap-python.png',
        bbox_inches = 'tight',
        transparent = True)
```



R

```{r}
#| label: topcategories

category_year <- expenditures_df %>%
  group_by(category) %>%
  summarise(top = sum(expenditure)) %>%
  arrange(desc(top)) %>%
  top_n(7)

category_year
```

```{r}
#| label: treemap
#| fig-width: 6.5

fig <- ggplot(category_year, aes(area = top, fill = category, label = category)) +
  geom_treemap() +
  geom_treemap_text(colour = "black", min.size = 0.3,
                    place = "centre", padding.x = grid::unit(3, "mm"), padding.y = grid::unit(3, "mm")) +
  theme(legend.position = "none") +
  scale_fill_brewer(palette = "Blues") +
  ggtitle("Treemap - R \nTop 7 Categories by Expenditure")

ggsave("images/treemap-r.png")
```

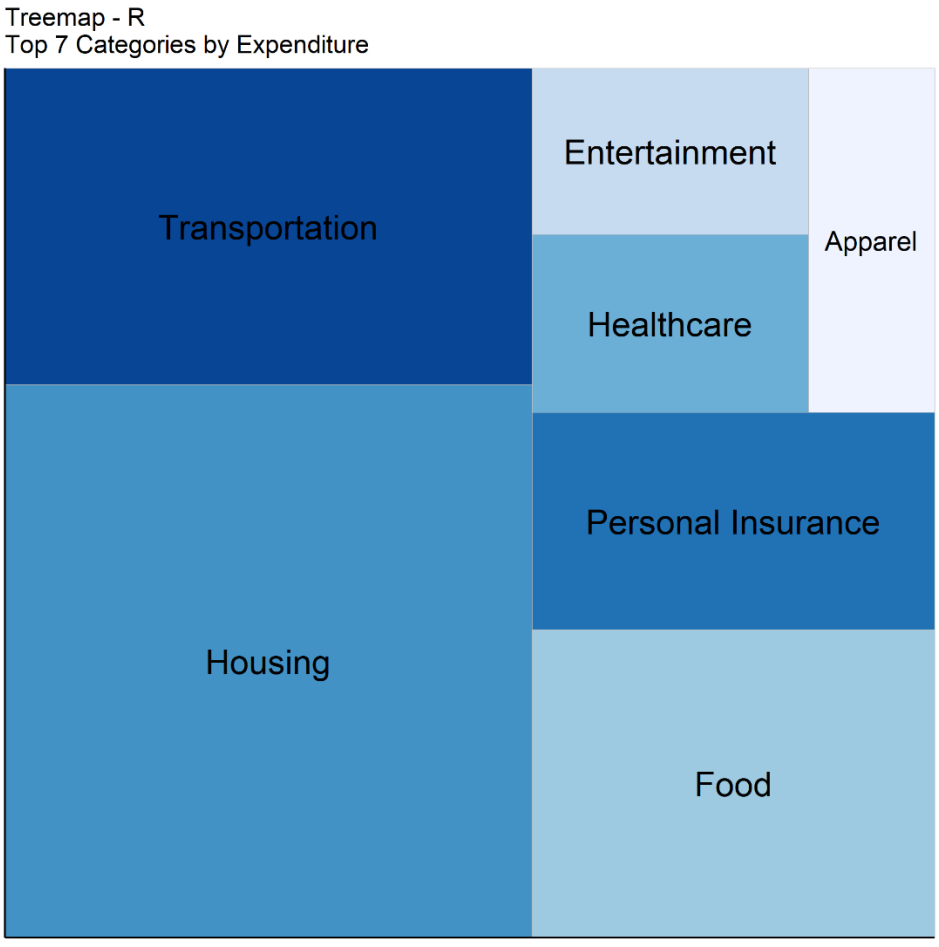
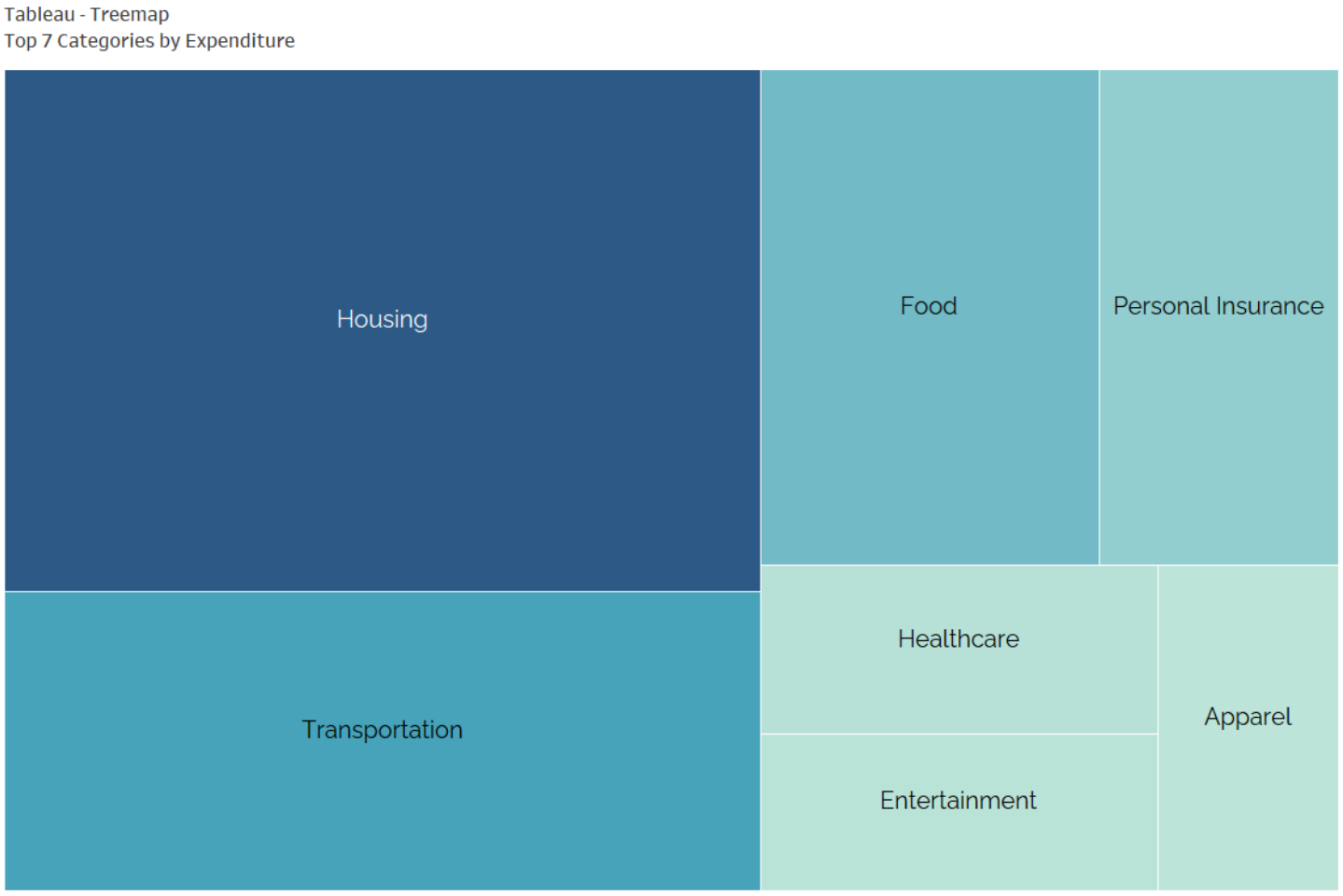


Tableau
(see Weeks5-6_Tableau.twb for code)



Area Chart

Python

```
# expenditures per year for sex = 1
expenditures_per_year = expenditures_df.groupby('year')
expenditures_sex1 = expenditures_per_year.apply(lambda x: x[x['sex'] == 1]['expenditure'].sum())
expenditures_sex1
```

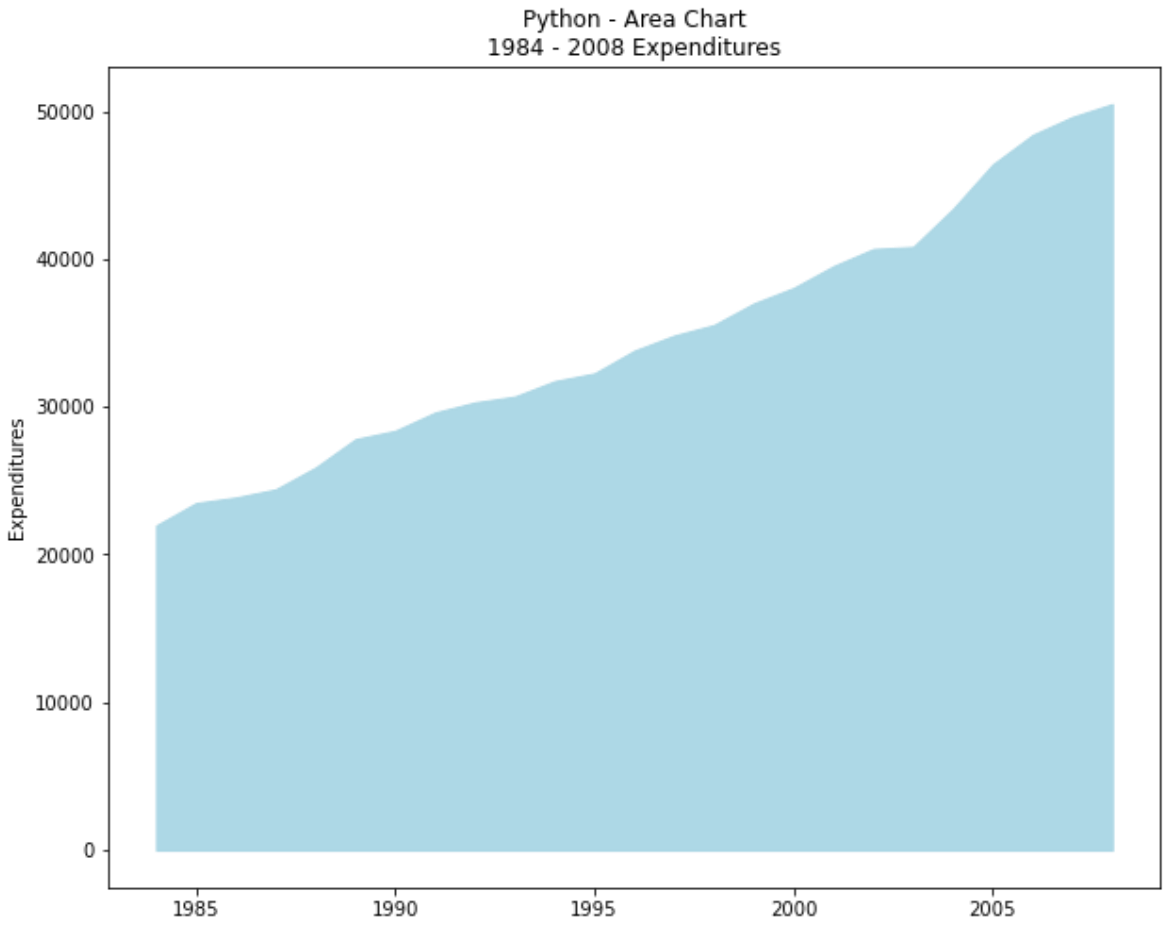
```
fig, ax = plt.subplots(figsize=(10,8))

ax.fill_between(expenditures_sex1.keys(), expenditures_sex1.values, color='lightblue')

plt.title('Python - Area Chart\n1984 - 2008 Expenditures')
plt.ylabel('Expenditures')
plt.xlabel('')

plt.show()

# Save figure
ax.get_figure().savefig('images/area-chart-python.png',
        bbox_inches = 'tight',
        transparent = True)
```



R

```{r}
#| label: expenditure per year for sex = 1

expenditures_per_year <- expenditures_df %>%
   group_by(year) %>%
   summarise_at(vars(expenditure), sum)

expenditures_per_year
```

```{r}
#| label: areachart
#| fig-width: 6.5

fig <- ggplot(expenditures_per_year, aes(x = year, y = expenditure)) +
   geom_area(fill = "lightblue") +
   ggtitle("Area Chart - R \n1984 - 2008 Expenditures") +
   xlab('') +
   ylab("Expenditures") +
   scale_x_continuous(breaks = c(1985, 1990, 1995, 2000, 2005))

ggsave("images/area-chart-r.png")
```

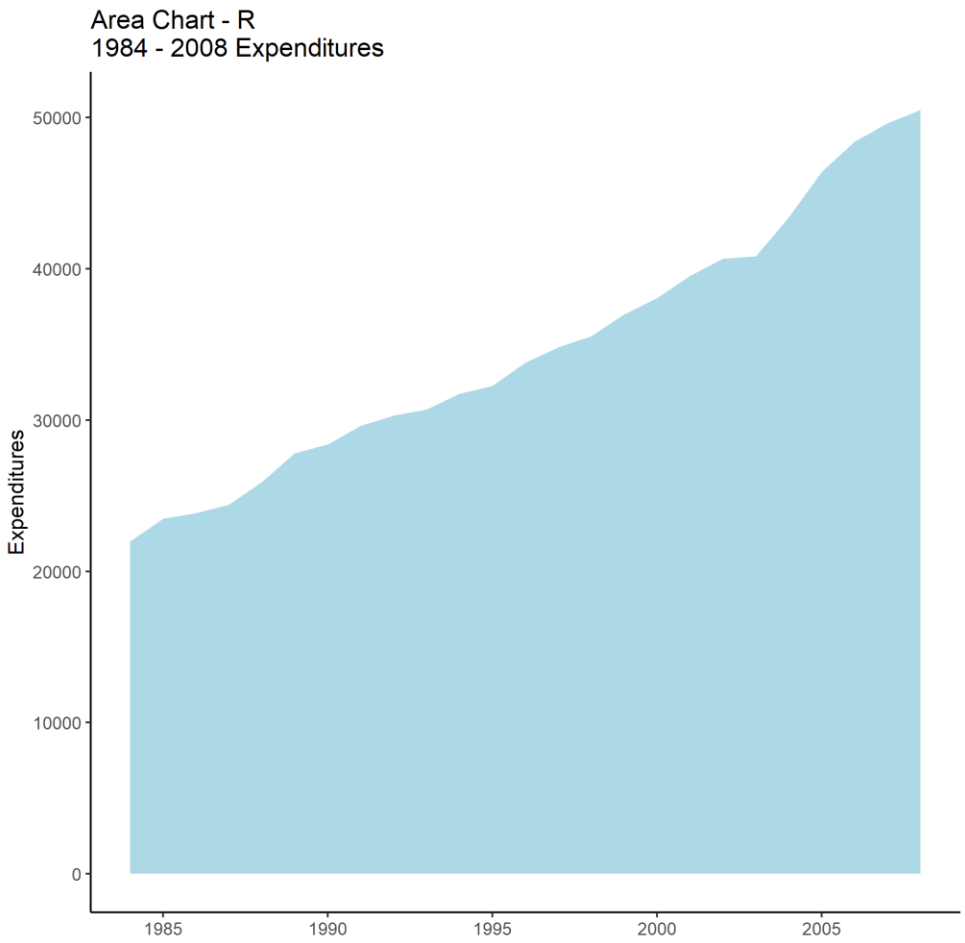
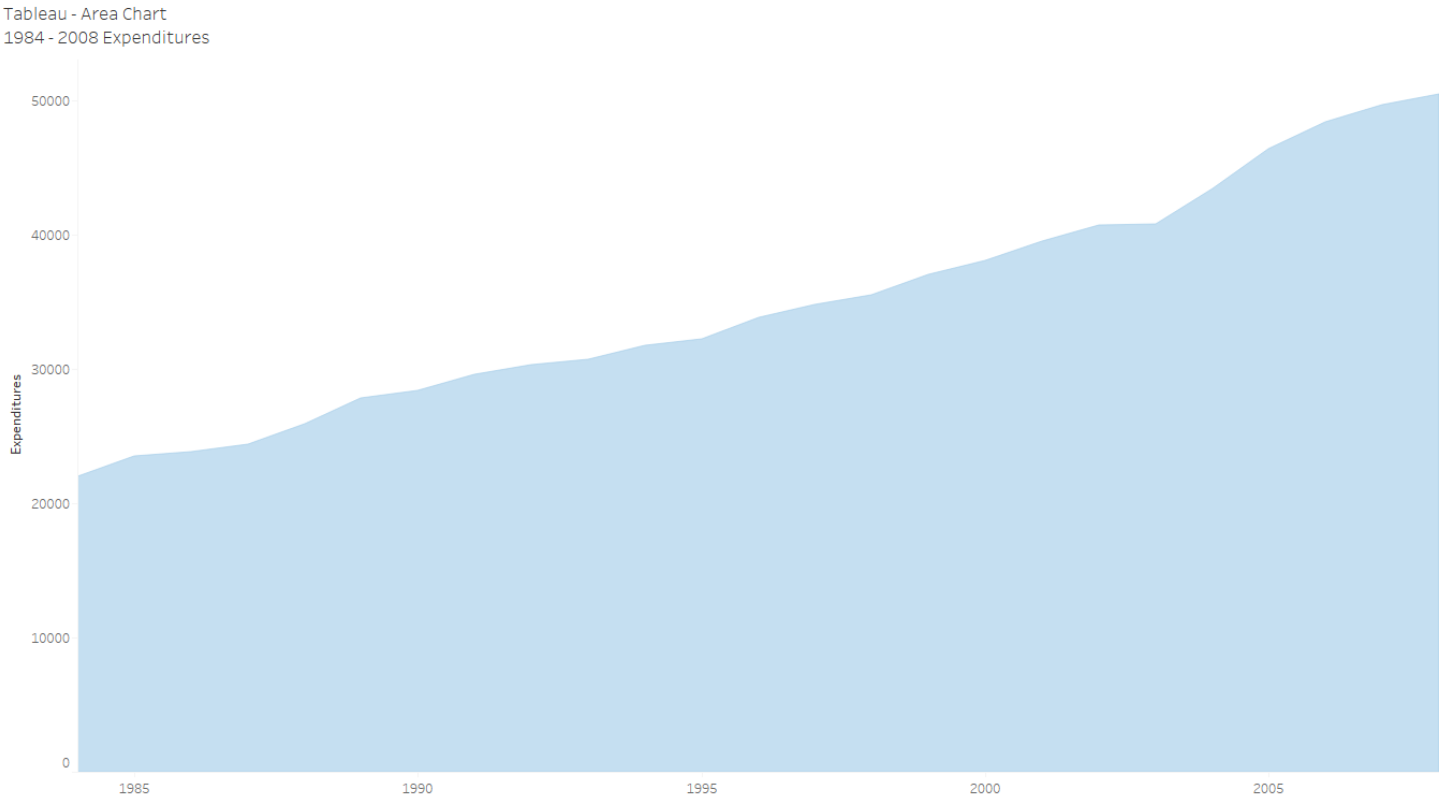


Tableau
(see Weeks5-6_Tableau.twb for code)



Stacked Area Chart

Python

```
year_2007 = list(unemployment_df[unemployment_df['Year'] == 2007]['Value'])
year_2008 = list(unemployment_df[unemployment_df['Year'] == 2008]['Value'])
year_2009 = list(unemployment_df[unemployment_df['Year'] == 2009]['Value'])

y_values = {
   '2009': year_2009,
   '2008': year_2008,
   '2007': year_2007
}

y_values
```

```
x = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec']
```

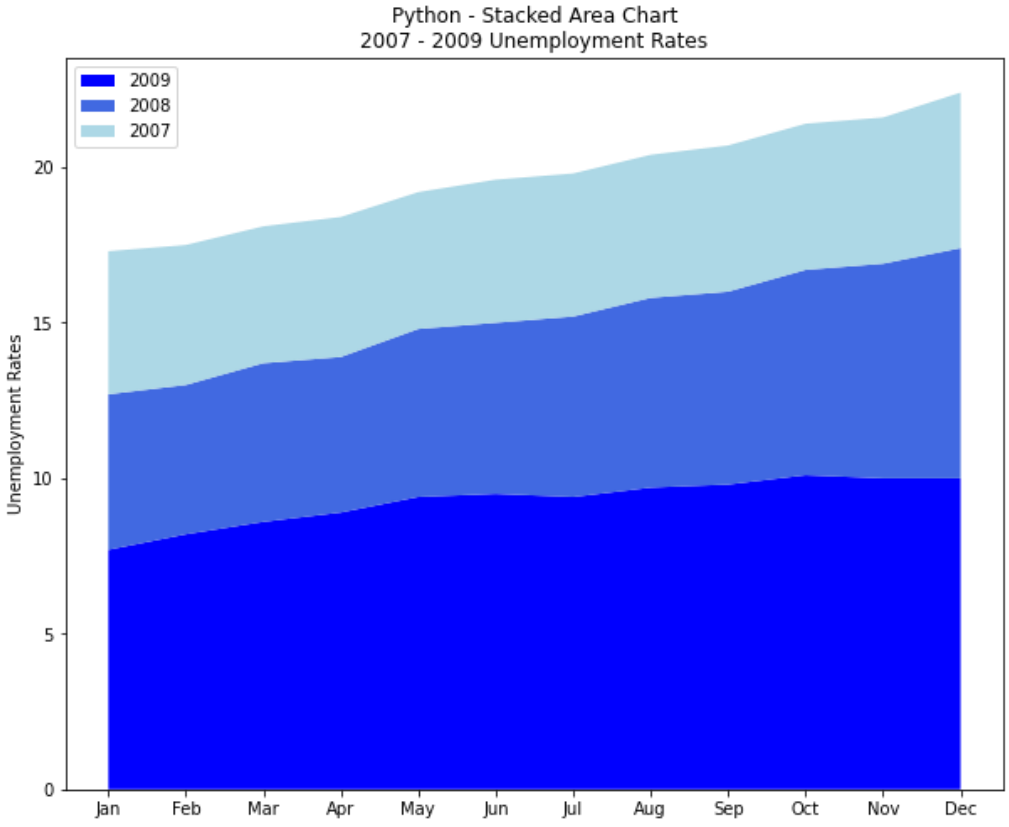
```
fig, ax = plt.subplots(figsize=(10,8))

ax.stackplot(x, y_values.values(), labels = y_values.keys(), colors = ['blue', 'royalblue', 'lightblue'])
plt.title('Python - Stacked Area Chart\n2007 - 2009 Unemployment Rates')
plt.ylabel('Unemployment Rates')
plt.xlabel('')

# legend
ax.legend(loc = 'upper left')

plt.show()

# Save figure
ax.get_figure().savefig('images/stacked-area-chart-python.png',
        bbox_inches = 'tight',
        transparent = True)
```



R

```{r}
#| label: yearcharacter

unemployment_df$Year <- as.character(unemployment_df$Year)
years <- c("2007", "2008", "2009")

three_years <- filter(unemployment_df, Year %in% years)

three_years
```

```{r}
#| label: stackedareachart
#| fig-width: 6.5

fig <- ggplot(three_years, aes(x = Period, y = Value, fill = Year, group = Year)) +
   geom_area(position = 'stack') +
   ggtitle("Stacked Area Chart - R \n2007 - 2009 Unemployment Rates") +
   xlab('') +
   scale_x_discrete(labels=c('Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul',
                             'Aug', 'Sep', 'Oct', 'Nov', 'Dec')) +
   ylab("Unemployment Rates") +
   scale_fill_manual(values = c('lightblue', 'royalblue', 'blue'))

ggsave("images/stacked-area-chart-r.png")
```

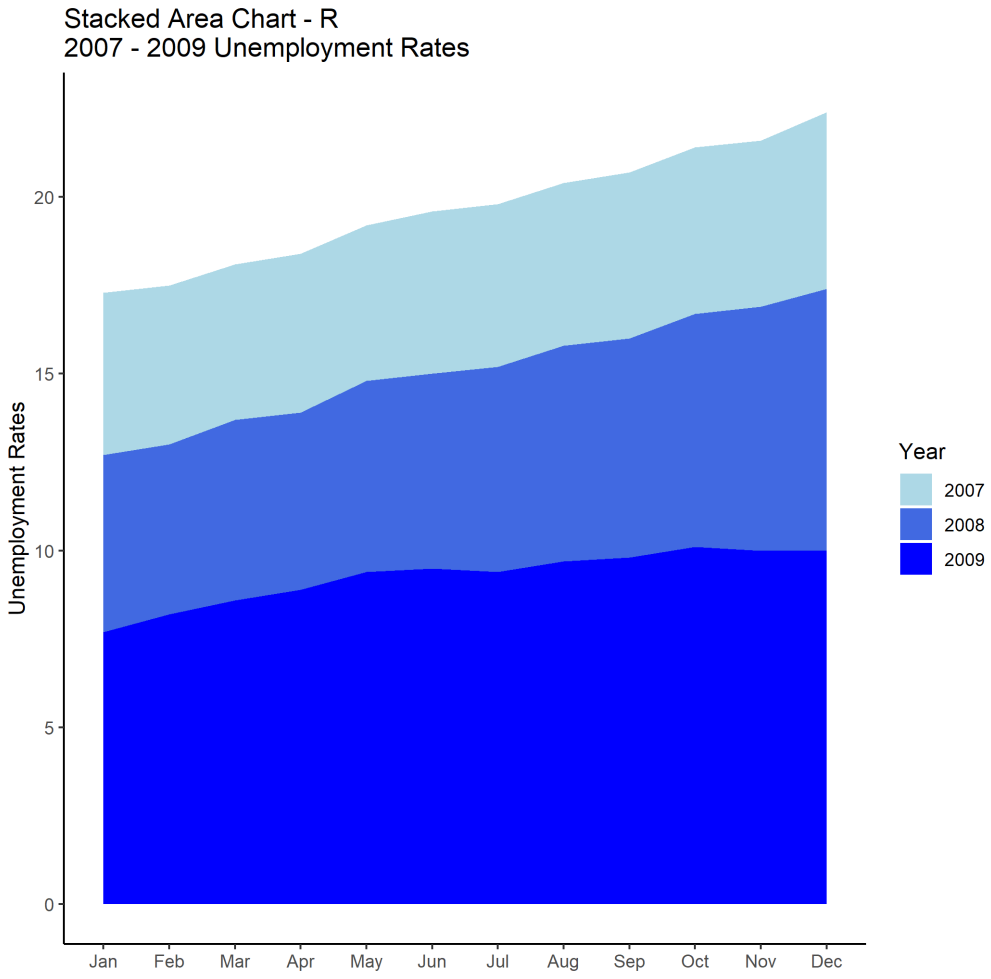


Tableau
(see Weeks5-6_Tableau.twb for code)