```
In [1]: import numpy as np
        import pandas as pd
        import re
        from math import pi

        from bokeh.io import output_file, show
        from bokeh.palettes import Category20c, Plasma256, plasma
        from bokeh.plotting import figure
        from bokeh.transform import cumsum
        from bokeh.models import LogColorMapper
        from bokeh.palettes import Viridis6 as palette
```

```
In [2]: # Load data
        cleaned = pd.read_csv('investments_with_dummies.csv')
        cleaned
```

Out[2]:

	Unnamed: 0	company_permalink	company_name	company_category_list	company_count
0	0	/organization/0-6-com	0-6.com	Curated Web	
1	1	/organization/004-technologies	004 Technologies	Software	
2	2	/organization/01games-technology	01Games Technology	Games	
3	3	/organization/0xdata	H2O.ai	Analytics	
4	4	/organization/0xdata	H2O.ai	Analytics	
5	5	/organization/0xdata	H2O.ai	Analytics	
6	6	/organization/0xdata	H2O.ai	Analytics	

```
In [3]:  # Beginning of Visualization for Q4

         # Drop all investments that did not occur in the United States
         stateInvestmentCount = cleaned.drop( cleaned[ cleaned['company_country_code
         stateInvestmentCount = stateInvestmentCount.drop(['Unnamed: 0', 'company_pe
         stateInvestmentCount = stateInvestmentCount.groupby('company_state_code').c
         stateInvestmentCount.drop('VI', inplace=True)
         stateInvestmentCount
```

Out[3]:

company_state_code	company_name	company_category_list	investor_name	raised_amount_usd	So
AK	4	4	4	3	
AL	66	66	66	48	
AR	90	90	90	75	
AZ	477	477	477	427	
CA	55620	55620	55619	50568	
CO	1912	1912	1912	1754	
CT	661	661	661	620	
DC	623	623	623	548	
DE	93	93	93	80	
FL	1228	1228	1228	1087	
GA	1378	1378	1378	1218	
HI	41	41	41	41	
IA	86	86	86	76	
ID	59	59	59	52	
IL	2311	2311	2311	2065	
IN	371	371	371	328	
KS	156	156	156	137	
KY	117	117	117	94	
LA	81	81	81	77	
MA	9628	9628	9627	9100	
MD	1293	1293	1293	1189	
ME	99	99	99	98	
MI	692	692	692	586	
MN	637	637	637	595	
MO	506	506	505	450	
MS	53	53	53	41	

company_state_code	company_name	company_category_list	investor_name	raised_amount_usd	So
MT	35	35	35	33	
NC	1331	1331	1331	1276	
ND	17	17	17	15	
NE	191	191	191	173	
NH	300	300	300	276	
NJ	1319	1319	1319	1212	
NM	213	213	213	201	
NV	307	307	307	254	
NY	13314	13314	13311	11915	
OH	1160	1160	1160	993	
OK	83	83	83	70	
OR	930	931	931	851	
PA	1861	1861	1861	1640	
RI	289	289	289	268	
SC	254	254	254	209	
SD	7	7	7	4	
TN	960	960	960	883	
TX	3902	3902	3902	3547	
UT	1032	1032	1032	936	
VA	1424	1424	1423	1229	
VT	85	85	85	60	
WA	3492	3492	3492	3244	
WI	395	395	395	374	
WV	12	12	12	7	
WY	6	6	6	5	

51 rows × 55 columns

```python
In [4]:  # Create bokeh figure

         from bokeh.sampledata.us_states import data as states

         output_file("investmentMap.html")

         # Reverse color scheme so darker is greater
         palette.reverse()

         state_tags = stateInvestmentCount.index.values

         # Remove Alaska and Hawaii for better map visualization
         state_tags = np.delete(state_tags, [0, 11])
         state_xs = [states[tag]["lons"] for tag in state_tags]
         state_ys = [states[tag]["lats"] for tag in state_tags]
         state_names = [states[tag]["name"] for tag in state_tags]

         investment_counts = stateInvestmentCount.loc[(stateInvestmentCount.index !=
         color_mapper = LogColorMapper(palette=palette)

         data=dict(
             x=state_xs,
             y=state_ys,
             name=state_names,
             rate=investment_counts,
         )

         TOOLS = "pan,wheel_zoom,reset,hover,save"

         p = figure(
             title="Distribution of Investments in US", plot_width = 1000, tools=TOO
             x_axis_location=None, y_axis_location=None,
             tooltips=[
                 ("Name", "@name"), ("Number of investments", "@rate"), ("(Long, Lat
             ])
         p.grid.grid_line_color = None
         p.hover.point_policy = "follow_mouse"

         p.patches('x', 'y', source=data,
                   fill_color={'field': 'rate', 'transform': color_mapper},
                   fill_alpha=0.7, line_color="white", line_width=0.5)

         show(p)
```

```
In [5]:  # Beginning of visualizations for Q1

         # Get possible sector tags from dataframe
         tags = cleaned.columns.values[-51:]
         investmentSums = []

         # For each tag, go through the dataframe of investments and add that invest
         # Then append that sum to a list of the investment sums for each of the sec
         for tag in tags:
             investmentSum = cleaned.loc[cleaned[tag] == 1, 'raised_amount_usd'].sum
             investmentSums.append(investmentSum)

         sectorInvestments = pd.Series({tag:total for (tag, total) in zip(tags, inve
         sectorInvestments = sectorInvestments.to_frame()
         sectorInvestments = sectorInvestments.reset_index()
         sectorInvestments = sectorInvestments.rename(columns={'index': 'Sectors', 0
         sectorInvestments = sectorInvestments.sort_values(by=['Total_Amount_Investe
         sectorInvestments
```

Out[5]:

	Sectors	Total_Amount_Invested
50	Other	1.201234e+12
2	Biotechnology	3.004740e+11
0	Software	2.542332e+11
3	E-Commerce	2.012236e+11
1	Mobile	1.876358e+11
4	Enterprise Software	1.449233e+11
7	Health Care	1.378122e+11
12	Internet	1.118102e+11
18	Clean Technology	1.060021e+11
13	Technology	9.471280e+10
6	Advertising	8.465768e+10
11	Finance	8.389714e+10
24	Semiconductors	7.666543e+10
9	Analytics	7.307581e+10
8	Social Media	7.053171e+10
5	Curated Web	6.568438e+10
10	SaaS	6.190084e+10
17	Hardware + Software	5.991142e+10
31	Web Hosting	5.419738e+10
32	Travel	5.005604e+10
20	Security	4.898736e+10

	Sectors	Total_Amount_Invested
23	Cloud Computing	4.897144e+10
19	Health and Wellness	4.727097e+10
14	Games	4.391113e+10
26	FinTech	4.302462e+10
30	Manufacturing	4.237196e+10
27	Services	4.226308e+10
15	Education	4.138003e+10
16	Apps	4.086637e+10
37	Retail	3.851323e+10
21	Big Data	3.717598e+10
34	Medical	3.478517e+10
43	Financial Services	3.343858e+10
28	Fashion	3.156079e+10
36	Networking	3.116324e+10
25	Video	3.073173e+10
22	Marketplaces	2.713639e+10
49	Entertainment	2.378811e+10
35	Messaging	2.237762e+10
42	Music	2.151589e+10
29	Search	2.131100e+10
48	Hospitality	2.058525e+10
45	Information Technology	1.853824e+10
33	Sales and Marketing	1.765681e+10
41	News	1.749029e+10
40	Digital Media	1.544376e+10
39	Media	1.458644e+10
44	Android	1.263644e+10
38	Startups	1.131488e+10
47	iPhone	1.047545e+10
46	Social Network Media	8.716398e+09

```
In [18]:  # Make bokeh figure of the sector breakdown of the top 20 sectors across al

          output_file("aggregateSectorInvestmentPie.html")

          data = sectorInvestments[:20].copy()
          data['angle'] = data['Total_Amount_Invested']/data['Total_Amount_Invested']
          data['color'] = Category20c[len(data)]

          print(data)

          p = figure(plot_height=600, title="Investment Distribution of All VCs", too
                     tools="hover", tooltips="@Sectors: @Total_Amount_Invested", x_ra

          p.wedge(x=0, y=1, radius=0.4,
                  start_angle=cumsum('angle', include_zero=True), end_angle=cumsum('a
                  line_color="white", fill_color='color', legend_field='Sectors', sou

          p.axis.axis_label=None
          p.axis.visible=False
          p.grid.grid_line_color = None

          show(p)
```

	Sectors	Total_Amount_Invested	angle	color
50	Other	1.201234e+12	2.206482	#3182bd
2	Biotechnology	3.004740e+11	0.551924	#6baed6
0	Software	2.542332e+11	0.466987	#9ecae1
3	E-Commerce	2.012236e+11	0.369617	#c6dbef
1	Mobile	1.876358e+11	0.344658	#e6550d
4	Enterprise Software	1.449233e+11	0.266202	#fd8d3c
7	Health Care	1.378122e+11	0.253140	#fdae6b
12	Internet	1.118102e+11	0.205378	#fdd0a2
18	Clean Technology	1.060021e+11	0.194710	#31a354
13	Technology	9.471280e+10	0.173973	#74c476
6	Advertising	8.465768e+10	0.155503	#a1d99b
11	Finance	8.389714e+10	0.154106	#c7e9c0
24	Semiconductors	7.666543e+10	0.140823	#756bb1
9	Analytics	7.307581e+10	0.134229	#9e9ac8
8	Social Media	7.053171e+10	0.129556	#bcbddc
5	Curated Web	6.568438e+10	0.120652	#dadaeb
10	SaaS	6.190084e+10	0.113702	#636363
17	Hardware + Software	5.991142e+10	0.110048	#969696
31	Web Hosting	5.419738e+10	0.099552	#bdbdbd
32	Travel	5.005604e+10	0.091945	#d9d9d9

```
In [7]: # Sort investors by amount invested
        ventureInvestments = cleaned.groupby(['investor_name']).sum()
        ventureInvestments = ventureInvestments.drop(columns=['Unnamed: 0'])
        topFifty = list(ventureInvestments.sort_values(by=['raised_amount_usd']).ta
        topFifty
        ventureInvestments.sort_values(by=['raised_amount_usd'])
```

Out[7]:

investor_name	raised_amount_usd	Software	Mobile	Biotechnology	E-Commerce	Enterprise Software	Curated Web
The Florida Technology Seed Capital Fund	0.000000e+00	0.0	0.0	0.0	0.0	0.0	0.0
Dato Dr Maimunah	0.000000e+00	0.0	1.0	0.0	0.0	0.0	0.0
Dato Sharil Tarmizi	0.000000e+00	0.0	1.0	0.0	0.0	0.0	0.0
Dato Syed Budriz Putra	0.000000e+00	0.0	1.0	0.0	0.0	0.0	0.0
Bharat Banka	0.000000e+00	0.0	0.0	0.0	0.0	0.0	0.0

```
In [8]:  # Isolate investments from top 50 VCs
         mask = []

         for row in cleaned['investor_name']:
             if row in topFifty:
                 mask.append(True)
             else:
                 mask.append(False)

         topFiftyInvestments = cleaned[mask]
         topFiftyCleanedInvestments = topFiftyInvestments.drop(['Unnamed: 0', 'compa
         topFiftyCleanedInvestments
```

Out[8]:

	company_name	company_category_list	investor_name	raised_amount_usd
20	One Inc.	Mobile	General Catalyst Partners	1000050.0
44	1000memories	Curated Web	Greylock Partners	2520000.0
91	1006.tv	Games|Media	Sequoia Capital	10000000.0
104	100Plus	Analytics	Founders Fund	500000.0
105	100Plus	Analytics	Greylock Partners	750000.0
106	1010data	Software	Norwest Venture Partners - NVP	35000000.0

```
In [9]:    # Get possible sector tags from dataframe
           tags = cleaned.columns.values[-51:]
           investmentSums = []

           # For each tag, go through the dataframe of investments and add that invest
           # Then append that sum to a list of the investment sums for each of the sec
           for tag in tags:
               investmentSum = topFiftyCleanedInvestments.loc[topFiftyCleanedInvestmen
               investmentSums.append(investmentSum)

           topFiftySectorInvestments = pd.Series({tag:total for (tag, total) in zip(ta
           topFiftySectorInvestments = topFiftySectorInvestments.to_frame()
           topFiftySectorInvestments = topFiftySectorInvestments.reset_index()
           topFiftySectorInvestments = topFiftySectorInvestments.rename(columns={'inde
           topFiftySectorInvestments = topFiftySectorInvestments.sort_values(by=['Tota
           topFiftySectorInvestments
```

Out[9]:

	Sectors	Total_Amount_Invested
50	Other	3.087448e+11
3	E-Commerce	5.781430e+10
0	Software	5.644804e+10
1	Mobile	5.156748e+10
12	Internet	4.386083e+10
4	Enterprise Software	3.898545e+10
13	Technology	2.843817e+10
24	Semiconductors	2.783208e+10
8	Social Media	2.608196e+10
6	Advertising	2.338961e+10
32	Travel	2.236253e+10
2	Biotechnology	2.042152e+10
11	Finance	1.967413e+10
9	Analytics	1.733629e+10
18	Clean Technology	1.732045e+10
5	Curated Web	1.708335e+10
20	Security	1.463007e+10
37	Retail	1.457577e+10
7	Health Care	1.377232e+10
10	SaaS	1.240407e+10
23	Cloud Computing	1.221800e+10
31	Web Hosting	1.110889e+10
14	Games	1.078425e+10

	Sectors	Total_Amount_Invested
30	Manufacturing	1.025086e+10
36	Networking	9.543591e+09
17	Hardware + Software	9.475200e+09
21	Big Data	9.300562e+09
15	Education	9.037512e+09
27	Services	8.415351e+09
16	Apps	8.385514e+09
19	Health and Wellness	8.096119e+09
26	FinTech	7.956654e+09
43	Financial Services	7.431280e+09
25	Video	6.928288e+09
35	Messaging	6.621850e+09
29	Search	6.293049e+09
28	Fashion	5.705732e+09
22	Marketplaces	5.298052e+09
49	Entertainment	4.903007e+09
45	Information Technology	4.468364e+09
41	News	4.009735e+09
42	Music	3.775806e+09
48	Hospitality	3.602285e+09
34	Medical	2.759994e+09
44	Android	2.701994e+09
40	Digital Media	2.671570e+09
33	Sales and Marketing	2.590871e+09
39	Media	2.491870e+09
46	Social Network Media	2.461864e+09
47	iPhone	2.104161e+09
38	Startups	2.010191e+09

```
In [10]: # Make bokeh figure of the sector breakdown of the top 20 sectors across al

         output_file("topFiftySectorPie.html")

         data = topFiftySectorInvestments[:20].copy()
         data['angle'] = data['Total_Amount_Invested']/data['Total_Amount_Invested']
         data['color'] = Category20c[len(data)]

         print(data)

         p = figure(plot_height=600, title="Investment Distribution of Top 50 VCs",
                    tools="hover", tooltips="@Sectors: @Total_Amount_Invested", x_ra

         p.wedge(x=0, y=1, radius=0.4,
                 start_angle=cumsum('angle', include_zero=True), end_angle=cumsum('a
                 line_color="white", fill_color='color', legend_field='Sectors', sou

         p.axis.axis_label=None
         p.axis.visible=False
         p.grid.grid_line_color = None

         show(p)
```

	Sectors	Total_Amount_Invested	angle	color
50	Other	3.087448e+11	2.329531	#3182bd
3	E-Commerce	5.781430e+10	0.436218	#6baed6
0	Software	5.644804e+10	0.425910	#9ecae1
1	Mobile	5.156748e+10	0.389085	#c6dbef
12	Internet	4.386083e+10	0.330937	#e6550d
4	Enterprise Software	3.898545e+10	0.294152	#fd8d3c
13	Technology	2.843817e+10	0.214571	#fdae6b
24	Semiconductors	2.783208e+10	0.209998	#fdd0a2
8	Social Media	2.608196e+10	0.196793	#31a354
6	Advertising	2.338961e+10	0.176479	#74c476
32	Travel	2.236253e+10	0.168729	#a1d99b
2	Biotechnology	2.042152e+10	0.154084	#c7e9c0
11	Finance	1.967413e+10	0.148445	#756bb1
9	Analytics	1.733629e+10	0.130805	#9e9ac8
18	Clean Technology	1.732045e+10	0.130686	#bcbddc
5	Curated Web	1.708335e+10	0.128897	#dadaeb
20	Security	1.463007e+10	0.110386	#636363
37	Retail	1.457577e+10	0.109977	#969696
7	Health Care	1.377232e+10	0.103914	#bdbdbd
10	SaaS	1.240407e+10	0.093591	#d9d9d9

```
In [11]:  # Beginning of Visualizations for Q5

          # Change all of the dates that the investments occured on to just the year
          def dateToYear(s):
              return s[0:4]

          cleanedYears = cleaned.copy()
          cleanedYears.loc[:, 'funded_at'] = cleanedYears['funded_at'].map(dateToYear

          cleanedYears = cleanedYears[(cleanedYears['funded_at'] != '1977') & (cleane
          years = sorted(cleanedYears['funded_at'].unique())
          years
```

Out[11]: ['1985',
          '1986',
          '1987',
          '1988',
          '1989',
          '1990',
          '1991',
          '1992',
          '1993',
          '1994',
          '1995',
          '1996',
          '1997',
          '1998',
          '1999',
          '2000',
          '2001',
          '2002',
          '2003',
          '2004',
          '2005',
          '2006',
          '2007',
          '2008',
          '2009',
          '2010',
          '2011',
          '2012',
          '2013',
          '2014',
          '2015']

```
In [12]:  tags = cleanedYears.columns.values[-51:]
          investmentSums = []
          yearlyInvestments = pd.DataFrame({tag: [] for tag in sorted(tags)})

          for year in years[10:]:
              print(year)
              for tag in tags:
                  investmentSum = cleanedYears.loc[(cleanedYears[tag] == 1) & (cleane
                  investmentSums.append(investmentSum)

              investmentPercentages = investmentSums

              sectorInvestmentsForYear = pd.Series({tag:total for (tag, total) in zip
              sectorInvestmentsForYear = sectorInvestmentsForYear.to_frame()
              sectorInvestmentsForYear = sectorInvestmentsForYear.reset_index()
              sectorInvestmentsForYear = sectorInvestmentsForYear.rename(columns={'in
              sectorInvestmentsForYear = sectorInvestmentsForYear.sort_values(by=['Se
              print(list(sectorInvestmentsForYear['Total_Amount_Invested']))
              yearlyInvestment = pd.DataFrame({sector:[investment] for (sector, inves
              print(yearlyInvestment)
              yearlyInvestments = yearlyInvestments.append(yearlyInvestment)
              investmentSums = []

          yearlyInvestments[['Other', 'Semiconductors']]
```

```
000.0, 15000000.0, 28000000.0, 14500000.0, 412834231.0, 24160000.0, 54800
0000.0, 0.0, 4780731266.417585, 125300000.0, 240000000.0, 0.0, 168000000.
0, 0.0, 19500000.0, 1196000000.0, 65000000.0, 0.0, 1852326056.417585, 0.
0, 271162868.417585, 2708000.0, 38470000.0, 1624600000.0, 0.0]
   Advertising  Analytics  Android       Apps  Big Data  Biotechnology  \
0  240000000.0  7529046.0      0.0  500000.0       0.0     45408740.0

   Clean Technology  Cloud Computing  Curated Web  Digital Media  ...  \
0               0.0      80000000.0  797018000.0    144000000.0  ...

        Services  Social Media  Social Network Media       Software    Startu
ps  \
0  1.196000e+09    65000000.0                   0.0  1.852326e+09
0.0

      Technology      Travel         Video  Web Hosting  iPhone
0  2.711629e+08  2708000.0  38470000.0  1.624600e+09     0.0

[1 rows x 51 columns]
```

```
In [13]: yearlyInvestments.index = years[10:]
         yearlyInvestments.loc[:,:] = yearlyInvestments.loc[:,:].div(yearlyInvestmen
         yearlyInvestments = yearlyInvestments * 100
         yearlyInvestments
```

Out[13]:

	Advertising	Analytics	Android	Apps	Big Data	Biotechnology	Clean Technology	Cloud Computing
1995	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1996	1.570722	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1997	2.996105	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1998	1.065583	0.000000	0.000000	9.412651	0.000000	0.568311	0.000000	0.000000
1999	1.431877	0.044919	0.000000	0.002983	0.000000	0.270916	0.000000	0.477292
2000	1.390537	0.017385	0.102778	1.912413	0.000000	1.266142	0.660719	0.358791
2001	0.714918	0.424854	0.058544	0.069603	0.000000	2.171440	0.100827	0.067243
2002	0.288719	0.492345	0.045687	0.047872	1.676500	7.410608	0.385754	1.605984
2003	11.010607	0.449859	0.000576	0.350107	0.115280	9.451693	0.025463	0.089665
2004	0.387829	0.206826	0.130883	0.214835	0.000000	16.772073	0.559932	0.182448
2005	1.161029	1.138177	0.314560	0.048188	0.081962	10.868195	0.777441	0.508497
2006	2.386618	1.107727	0.113898	0.344760	0.272258	10.596486	2.027198	0.336554
2007	2.401473	1.476678	0.137857	0.297570	0.194560	14.151119	4.021017	0.874870
2008	2.421692	1.407794	0.138225	0.272626	0.388453	6.945444	9.155554	0.623789
2009	1.767986	0.981466	0.101222	0.359827	0.213211	14.000923	5.152028	0.522079
2010	3.099854	1.851699	0.177446	0.391639	0.440720	12.928977	6.942664	0.760069
2011	4.650398	1.397023	0.459500	0.783893	0.528806	5.597586	3.678688	1.047500
2012	2.472634	2.530945	0.288681	1.163726	1.030487	7.089195	2.994726	2.061224
2013	1.644857	2.443368	0.497404	1.144384	1.376225	6.428236	2.240725	1.726551
2014	1.074226	2.081659	0.433307	1.192155	1.418563	4.346664	1.462547	1.151702
2015	0.875664	1.786742	0.270793	1.493163	1.105047	4.585618	0.670686	1.416817

21 rows × 51 columns

```
In [14]:  # Making Bokeh figure of Percentage Bar Graph of the top 50 Sectors in term

          from bokeh.core.properties import value
          from bokeh.io import show, output_file
          from bokeh.plotting import figure

          output_file("percentageBarGraphTop50.html")

          colors = Category20c[20]*3
          colors = colors[:51]

          data = yearlyInvestments

          p = figure(x_range=data.index.values, plot_height=1000, plot_width=1500, ti
                     toolbar_location=None, tools="hover", tooltips="$name: @$name %"

          p.vbar_stack(tags, x='index', width=0.9, color=colors, source=data)

          p.y_range.start = 0
          p.x_range.range_padding = 0.1
          p.xgrid.grid_line_color = None
          p.axis.minor_tick_line_color = None
          p.outline_line_color = None

          show(p)
```

```
In [15]:  tags = cleanedYears.drop(columns = ['Other']).columns.values[-50:]
          investmentSums = []
          yearlyInvestments = pd.DataFrame({tag: [] for tag in sorted(tags)})

          for year in years[10:]:
              print(year)
              for tag in tags:
                  investmentSum = cleanedYears.loc[(cleanedYears[tag] == 1) & (cleane
                  investmentSums.append(investmentSum)

              investmentPercentages = investmentSums

              sectorInvestmentsForYear = pd.Series({tag:total for (tag, total) in zip
              sectorInvestmentsForYear = sectorInvestmentsForYear.to_frame()
              sectorInvestmentsForYear = sectorInvestmentsForYear.reset_index()
              sectorInvestmentsForYear = sectorInvestmentsForYear.rename(columns={'in
              sectorInvestmentsForYear = sectorInvestmentsForYear.sort_values(by=['Se
              print(list(sectorInvestmentsForYear['Total_Amount_Invested']))
              yearlyInvestment = pd.DataFrame({sector:[investment] for (sector, inves
              print(yearlyInvestment)
              yearlyInvestments = yearlyInvestments.append(yearlyInvestment)
              investmentSums = []

          yearlyInvestments



          yearlyInvestments.index = years[10:]

          yearlyInvestments.loc[:,:] = yearlyInvestments.loc[:,:].div(yearlyInvestmen
          yearlyInvestments = yearlyInvestments * 100
          yearlyInvestments
```

```
1995
[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8000000.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 24400000.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 8000000.0, 0.0, 0.0, 16400000.0,
0.0, 0.0, 0.0, 0.0, 0.0, 8000000.0, 0.0, 16400000.0, 0.0, 0.0, 2000000.0,
0.0]
   Advertising  Analytics  Android  Apps  Big Data  Biotechnology  \
0          0.0        0.0      0.0   0.0       0.0            0.0

   Clean Technology  Cloud Computing  Curated Web  Digital Media  ...  \
0               0.0              0.0          0.0            0.0  ...

   Services  Social Media  Social Network Media  Software  Startups  \
0       0.0           0.0                   0.0  8000000.0       0.0

   Technology  Travel  Video  Web Hosting  iPhone
0  16400000.0     0.0    0.0    2000000.0     0.0

[1 rows x 50 columns]
```

```python
# Making Bokeh figure of Percentage Bar Graph of the top 50 Sectors (withou

from bokeh.core.properties import value
from bokeh.io import show, output_file
from bokeh.plotting import figure

output_file("percentageBarGraphTop50WithoutOther.html")

colors = Category20c[20]*3
colors = colors[:50]

data = yearlyInvestments

p = figure(x_range=data.index.values, plot_height=1000, plot_width=1500, ti
           toolbar_location=None, tools="hover", tooltips="$name: @$name %"

p.vbar_stack(tags, x='index', width=0.9, color=colors, source=data)

p.y_range.start = 0
p.x_range.range_padding = 0.1
p.xgrid.grid_line_color = None
p.axis.minor_tick_line_color = None
p.outline_line_color = None

show(p)
```