Configuration

In [ ]:
```
args = {
  'country': 'united_states',
  'start_year': 1960,
  'end_year': 2020, # exclusive
  'real_gdp_scale': 1000000,
  'bubble_scale': 0.5,
  'bubble_min_count': 100,
}
```

Libraries

In [ ]:
```
import pandas as pd
import importlib

import constants.tags as IndustryConstants
import utils.file as File
import utils.data as DataUtils
import utils.visualiser as Visualiser

import utils.firm as FirmUtils
import utils.investor as InvestorUtils

STEM_stack_labels = ['Not STEM', 'STEM']
STEM_stack_colors = ['green', 'lightgreen']
public_funded_stack_labels = ['Private Funded', 'Public Funded']
public_funded_stack_colors = ['blue', 'lightblue']

def get_scaled_year_count(year_count, real_gdp):
  scaled_year_count = year_count / real_gdp.loc[year_count.index] * args['re

  return scaled_year_count

def get_STEM_public_year_percent(firms):
  STEM_public_year_percent = DataUtils.get_grouped_percent(firms, ['Founded
    .unstack() \
    .reset_index() \
    .set_index('Founded Year')
  return STEM_public_year_percent

def visualise_bubble(industry_firms, industry, industry_year_count, fed_rate
  industry_public_year_count = DataUtils.get_public_year_count(industry_firm

  industry_bubble = {
    'All': industry_year_count,
    'Private Funded': industry_public_year_count['Private Funded'],
    'Public Funded': industry_public_year_count['Public Funded'],
    'Interest Rate': fed_rate,
  }

  Visualiser.plot_dict(industry_bubble, f"Bubble: {industry}", '')
```

```
def visualise_industry(firms, industry, fed_rate):
  industry_firms = DataUtils.filter_industry(firms, industry)
  industry_year_count = DataUtils.get_year_count(industry_firms)

  industry_max_count = industry_year_count.max()
  industry_current_count = industry_max_count
  for year in range(args['end_year'], 0, -1):
    if year in industry_year_count:
      industry_current_count = industry_year_count[year]
      break
  industry_bubble_scale = industry_current_count / industry_max_count

  if industry_max_count > args['bubble_min_count'] and args['bubble_scale']
    visualise_bubble(industry_firms, industry, industry_year_count, fed_rate

def visualise_industry_group(firms, industry_group, fed_rate):
  industry_group_firms = DataUtils.filter_industry_group(firms, industry_gro

  industry_group_year_count = DataUtils.get_year_count(industry_group_firms)
  industry_group_public_year_count = DataUtils.get_public_year_count(industr
  industry_group_public_year_percent = DataUtils.get_public_year_percent(ind

  Visualiser.plot(industry_group_year_count, f"{industry_group}: Firms Found
  Visualiser.plot_dict(industry_group_public_year_count, f"{industry_group}:
  Visualiser.stack(industry_group_public_year_percent, f"{industry_group}: P

  industry_group_industries = DataUtils.get_industry_group_industries(indust
  for industry in industry_group_industries:
    visualise_industry(firms, industry, fed_rate)
```

Load Data

In [ ]:
```
# Read Company Data
firms = File.read_pickle(f"{args['country']}_firms")
```

In [ ]:
```
# Read Investor Data
investors = File.read_pickle(f"investors")
```

In [ ]:
```
# Get Public Investors
public_investor_list = InvestorUtils.get_public_investors(investors).tolist(
```

In [ ]:
```
# Read Domain Data
domain_created_year_map = File.read_pickle('domain_created_year_map')
```

In [ ]:
```
# Update Company Data
firms = FirmUtils.enrich_founded_year(firms, domain_created_year_map)
firms = FirmUtils.enrich_stem(firms, IndustryConstants.stem_tags)
firms = FirmUtils.enrich_public_funded(firms, public_investor_list)
firms = firms[firms['Company Type'] == 'For Profit']
firms = firms[pd.notna(firms['Founded Year'])]
firms = firms[firms['Founded Year'] >= args['start_year']]
firms = firms[firms['Founded Year'] < args['end_year']]
firms = firms.reset_index(drop=True)
```

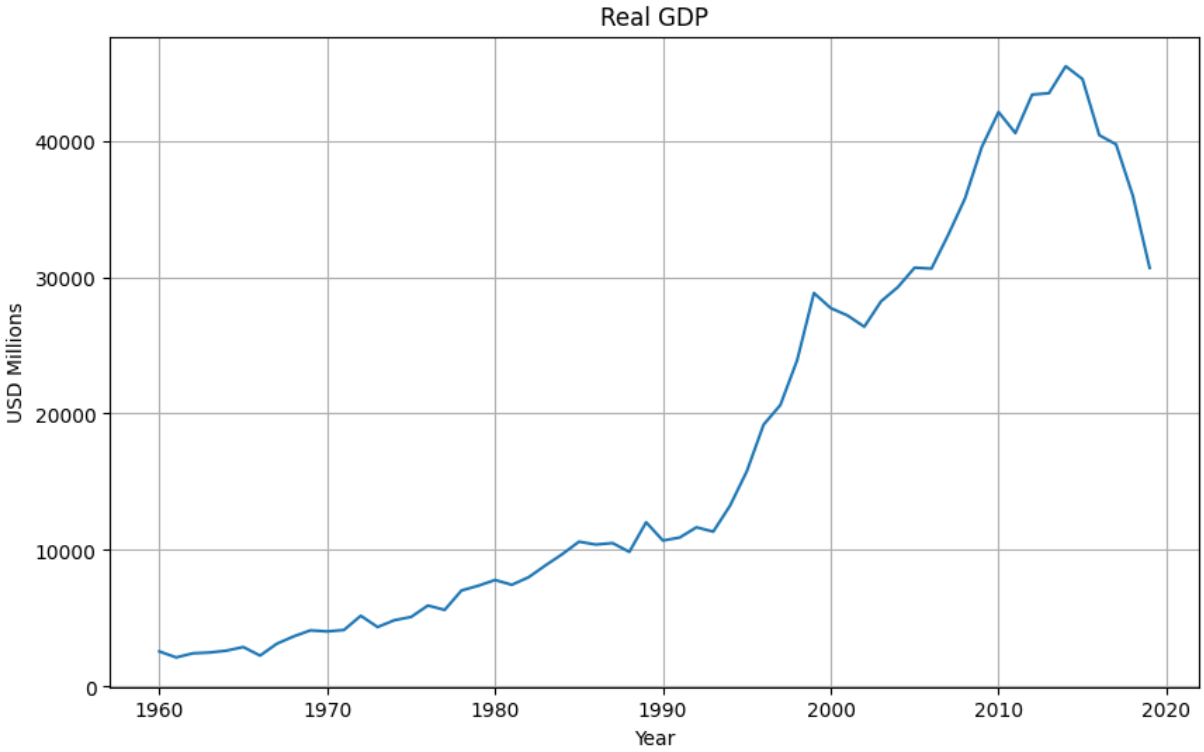
In [ ]:
```
# Read Macro Data
real_gdp = File.read_pickle('real_gdp')
fed_rate = File.read_pickle('fed_rate')
```

In [ ]:
```
# Update Macro Data
real_gdp = real_gdp['United States']
real_gdp = real_gdp[real_gdp.index >= args['start_year']]
real_gdp = real_gdp[real_gdp.index < args['end_year']]

fed_rate = fed_rate[fed_rate.index >= args['start_year']]
fed_rate = fed_rate[fed_rate.index < args['end_year']]
```

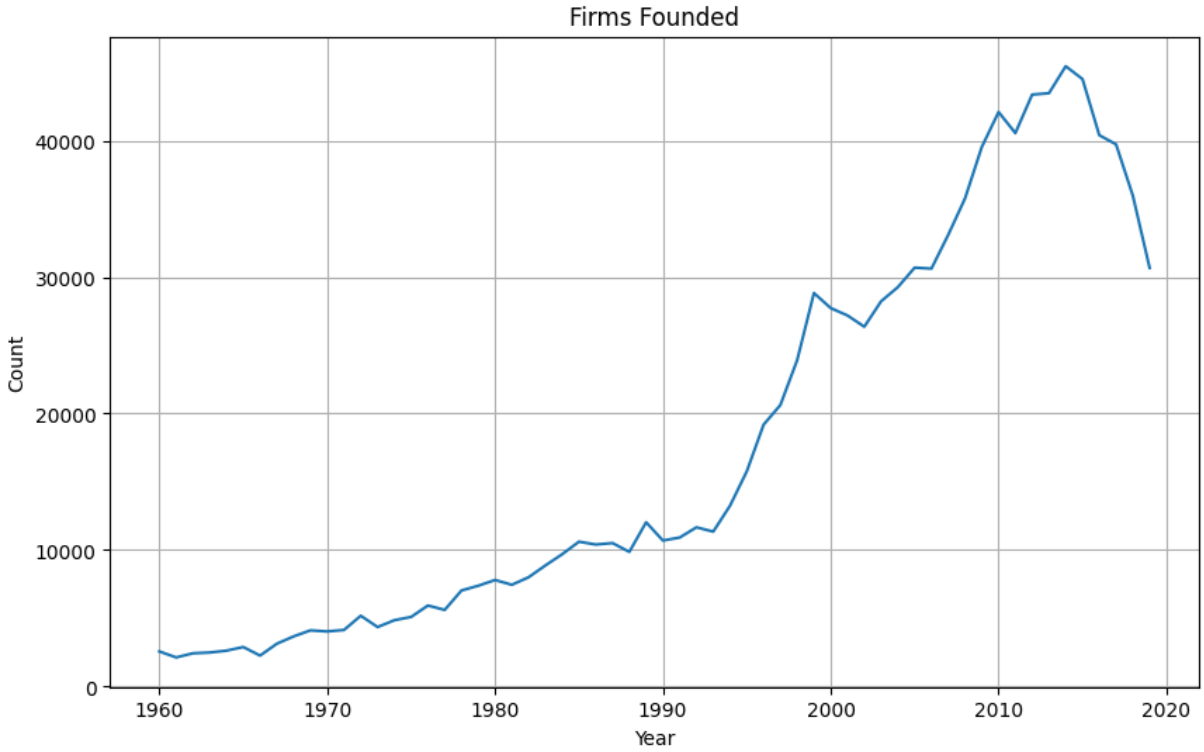
Broad Facts

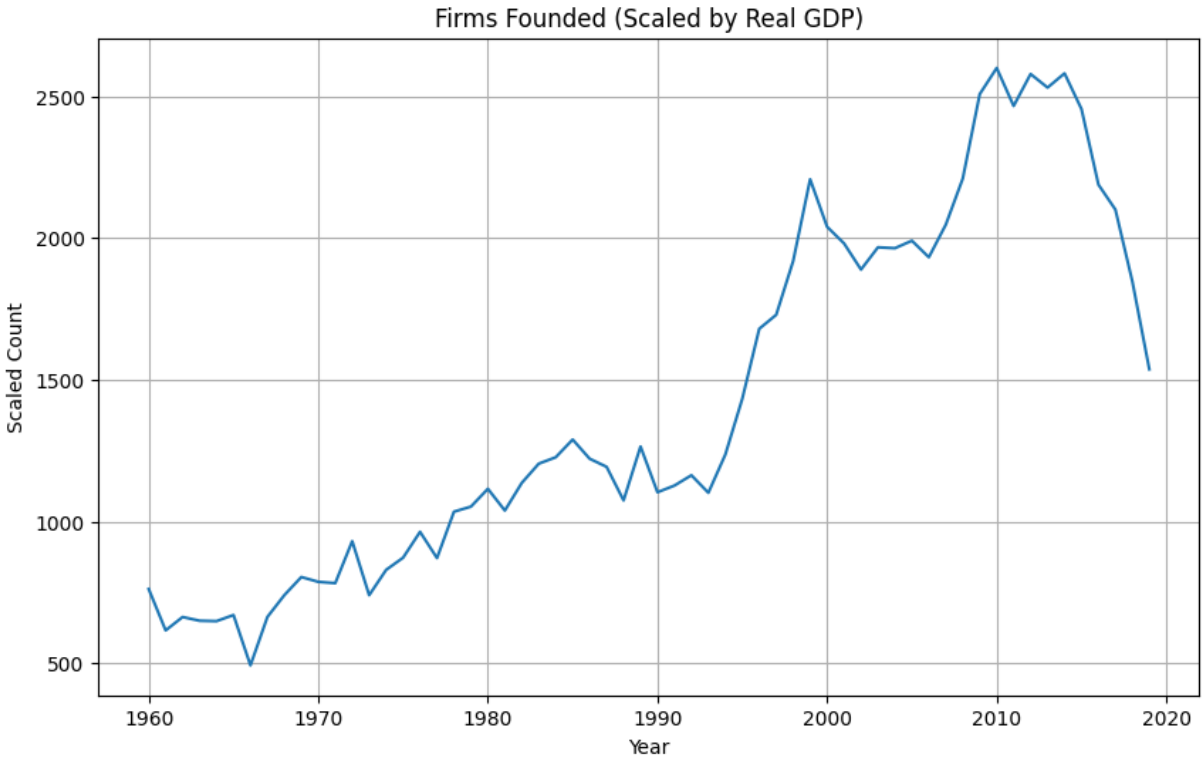
In [ ]:
```
year_count = DataUtils.get_year_count(firms)
scaled_year_count = get_scaled_year_count(year_count, real_gdp)

STEM_year_percent = DataUtils.get_STEM_year_percent(firms)
public_year_percent = DataUtils.get_public_year_percent(firms)
STEM_public_year_percent = get_STEM_public_year_percent(firms)

# To Do, get recession yars
```

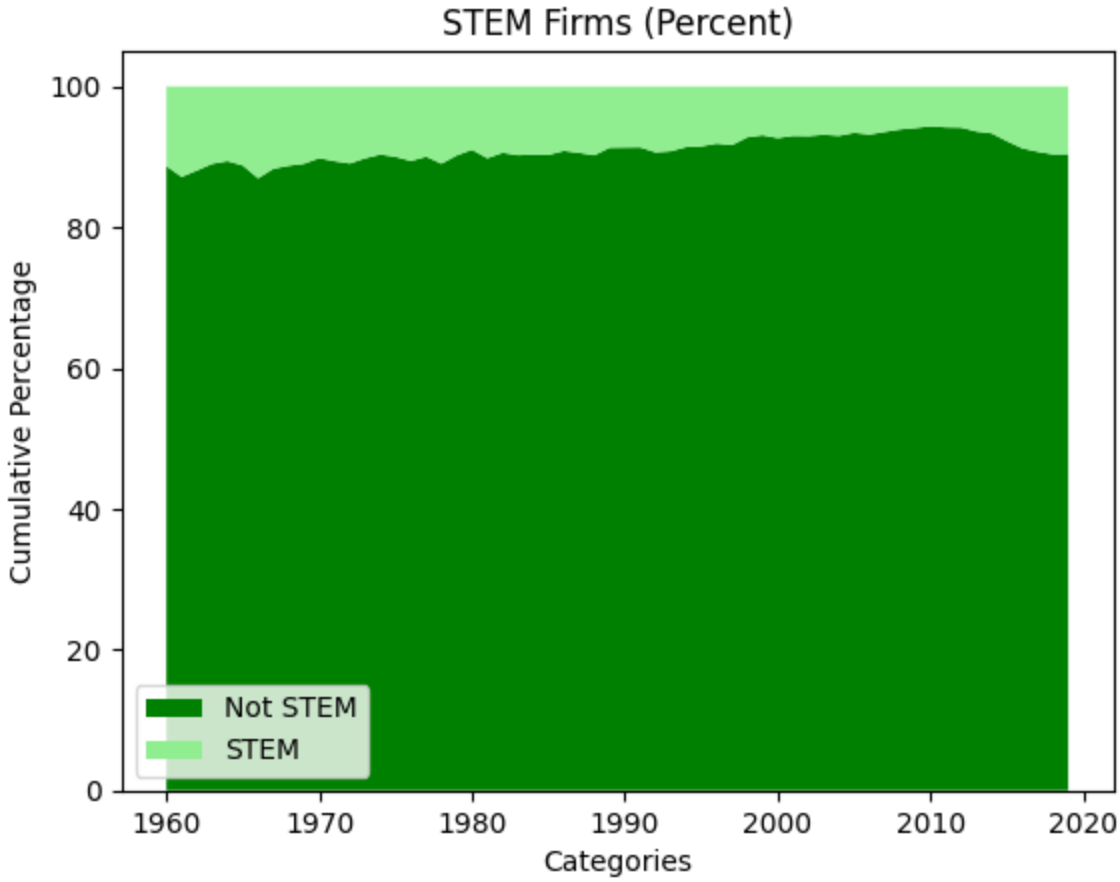
In [ ]:
```
Visualiser.plot(year_count, 'Real GDP', 'USD Millions')
```

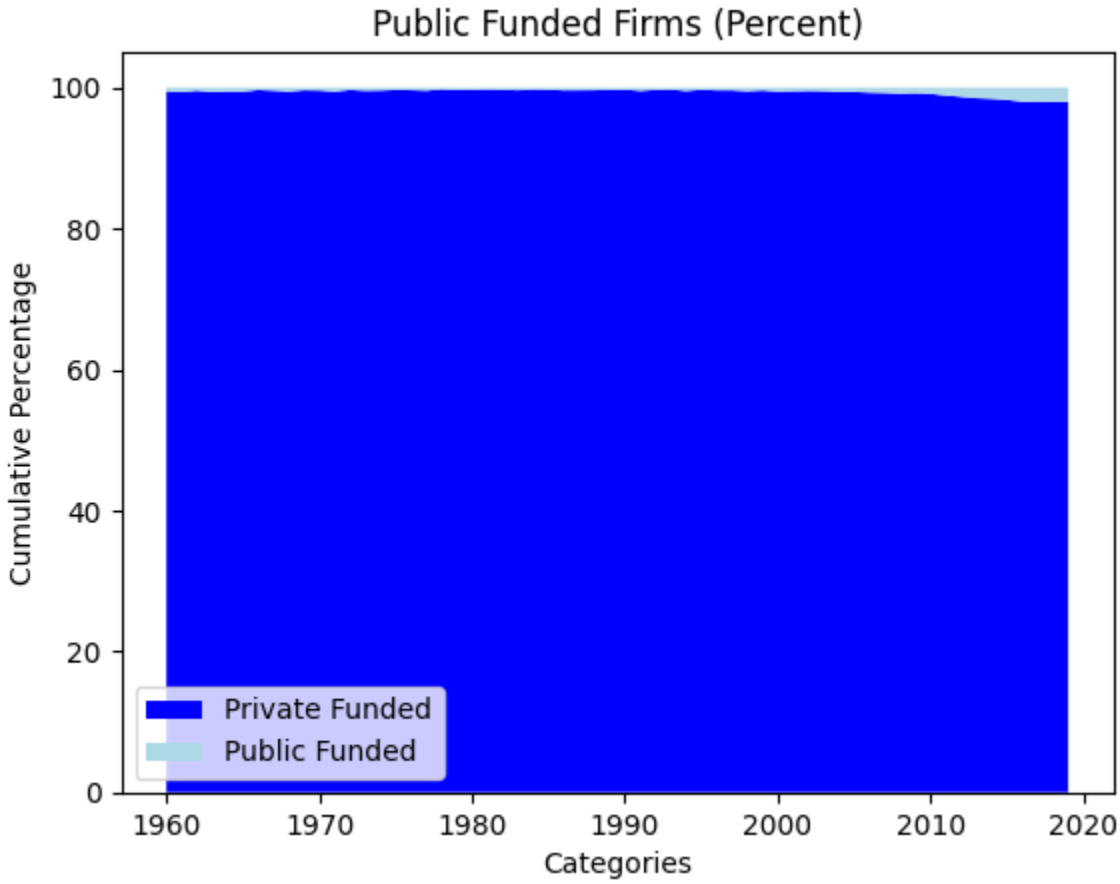


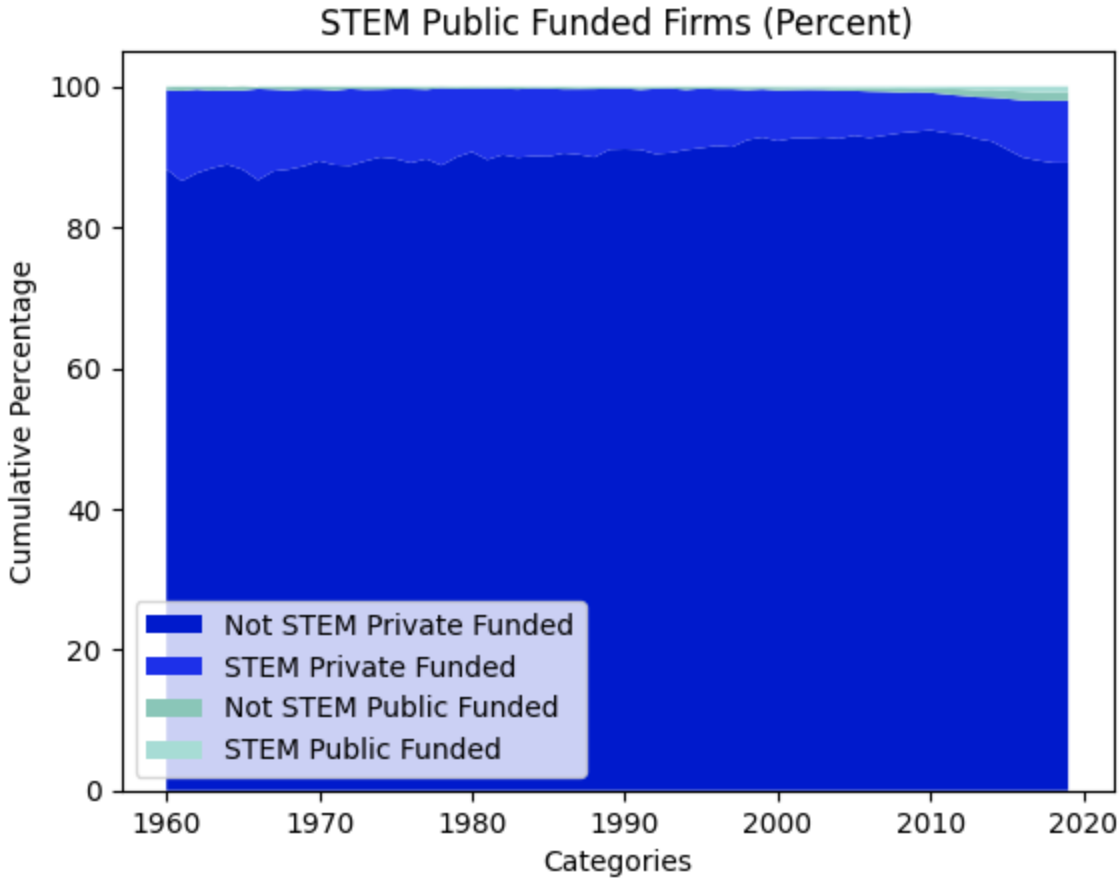
In [ ]:
```
# To Do, add interest rate

Visualiser.plot(year_count, 'Firms Founded', 'Count')
Visualiser.plot(scaled_year_count, 'Firms Founded (Scaled by Real GDP)', 'Sc
Visualiser.stack(STEM_year_percent, 'STEM Firms (Percent)', STEM_stack_label
Visualiser.stack(public_year_percent, 'Public Funded Firms (Percent)', publi

STEM_public_year_percent_colors = [Visualiser.mid_color('green','blue', 0.8)

Visualiser.stack(STEM_public_year_percent, 'STEM Public Funded Firms (Percen
                 ['Not STEM Private Funded', 'STEM Private Funded', 'Not STE
```


Firms Founded
Count
Year
0
10000
20000
30000
40000
1960
1970
1980
1990
2000
2010
2020


Firms Founded (Scaled by Real GDP)
Scaled Count
Year
500
1000
1500
2000
2500
1960
1970
1980
1990
2000
2010
2020

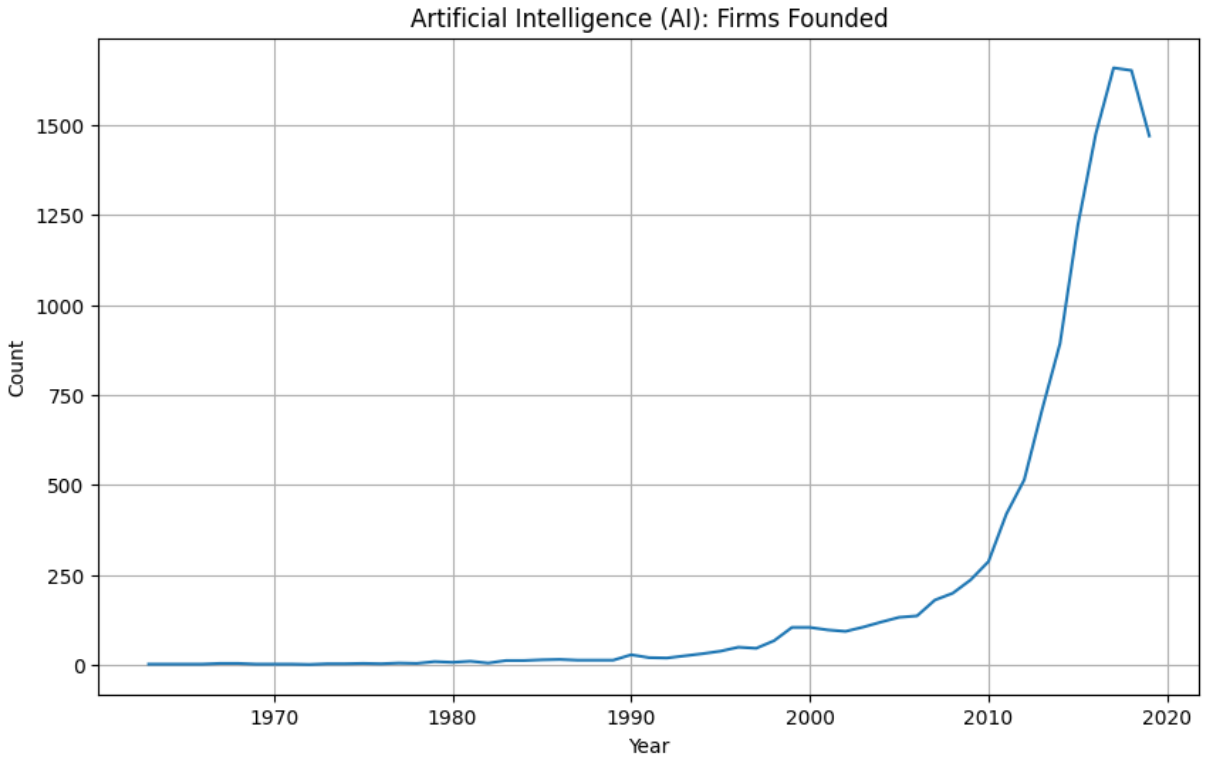

STEM Firms (Percent)
Cumulative Percentage
Categories
Not STEM
STEM
0
20
40
60
80
100
1960
1970
1980
1990
2000
2010
2020

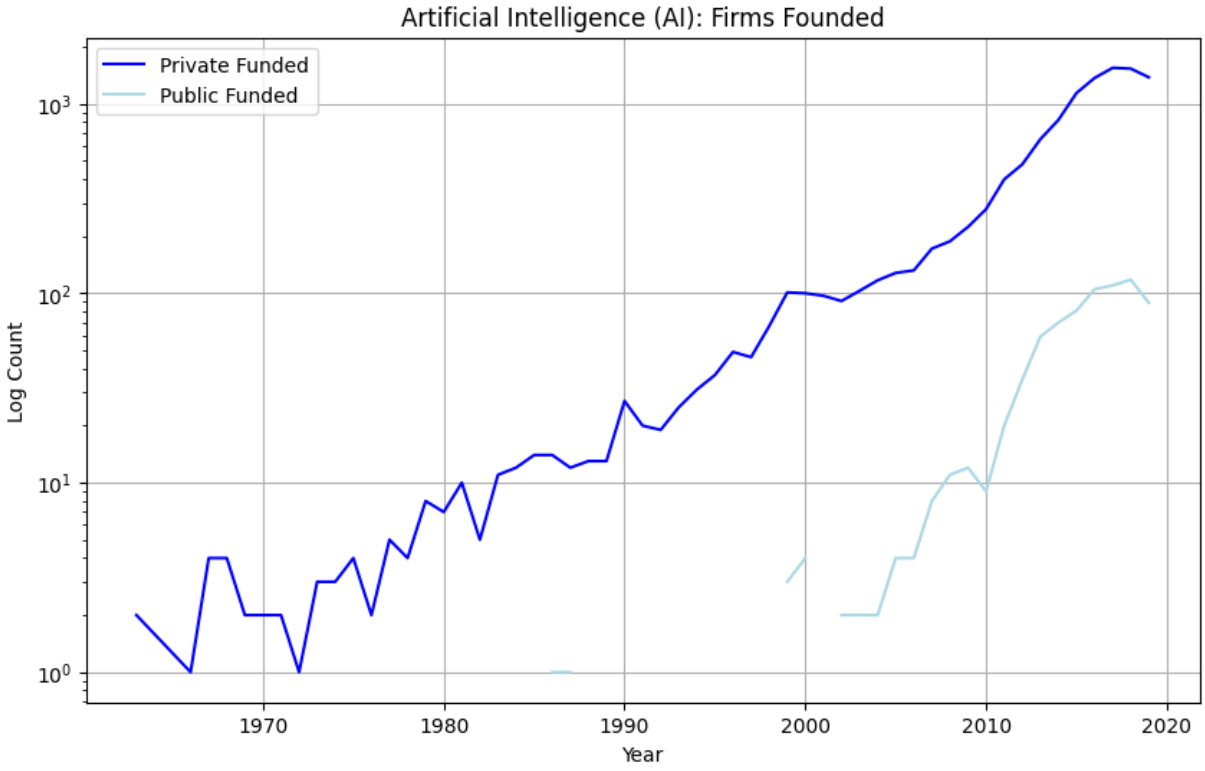

Public Funded Firms (Percent)
Cumulative Percentage
Categories
Private Funded
Public Funded
0
20
40
60
80
100
1960
1970
1980
1990
2000
2010
2020

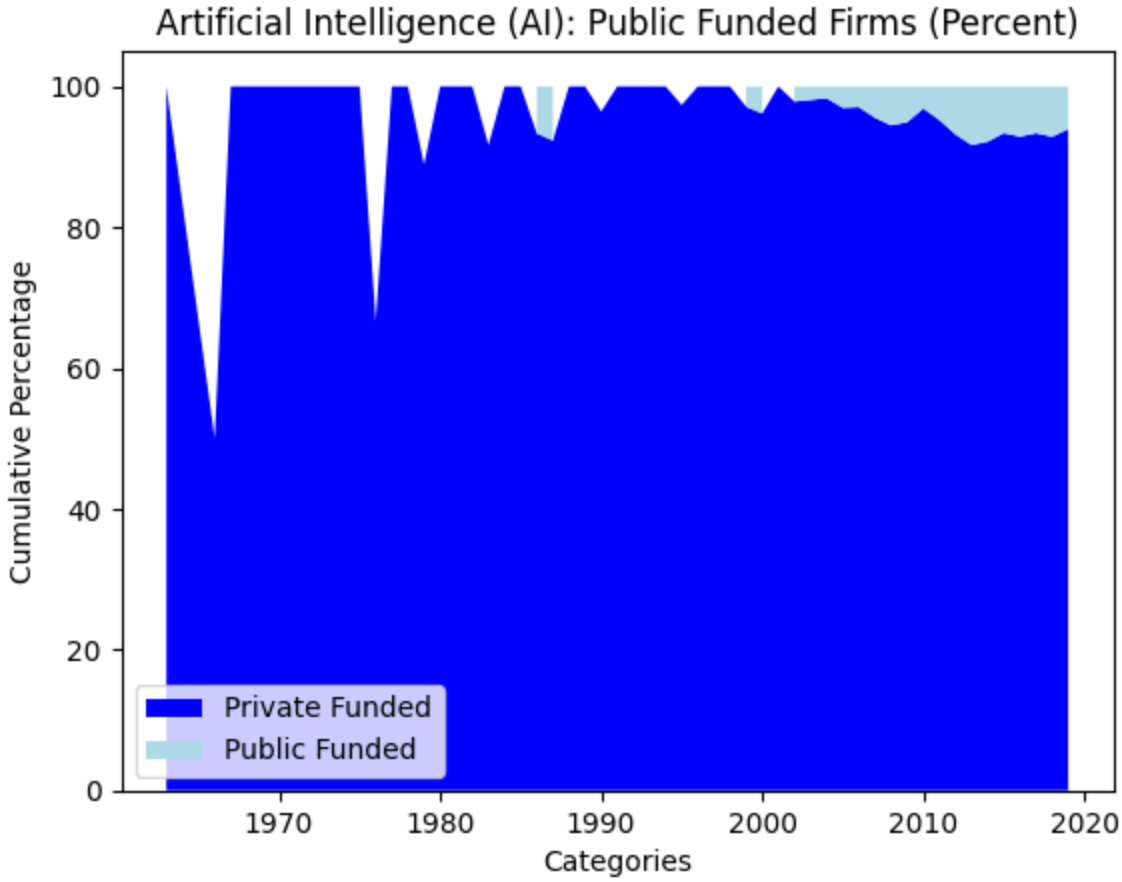


In [ ]:
```
# industries = DataUtils.get_industries()

# bubble_industries = []
# industry_year_count_map = {}
# industry_bubble_scale_map = {}

# for industry in industries:
#   industry_firms = DataUtils.filter_industry(firms, industry)
#   industry_year_count = DataUtils.get_year_count(industry_firms)
#   industry_max_count = industry_year_count.max()

#   industry_year_count_map[industry] = industry_year_count

#   industry_current_count = industry_max_count
#   for year in range(args['end_year'], 0, -1):
#     if year in industry_year_count:
#         industry_current_count = industry_year_count[year]
#         break

#   industry_bubble_scale = industry_current_count / industry_max_count
#   industry_bubble_scale_map[industry] = industry_bubble_scale

#   if industry_max_count > args['bubble_min_count'] and args['bubble_scale'
#     bubble_industries.append(industry)
```

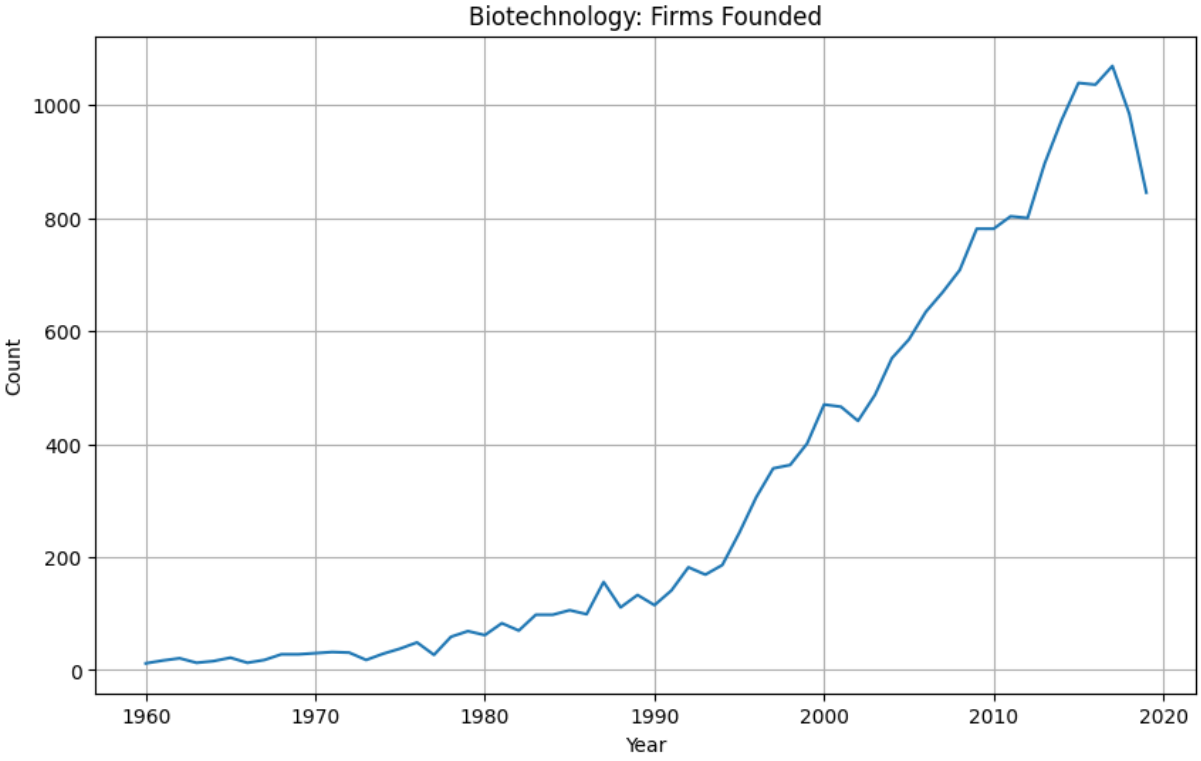
In [ ]:
```
visualise_industry_group(firms, 'Artificial Intelligence (AI)', fed_rate)
```

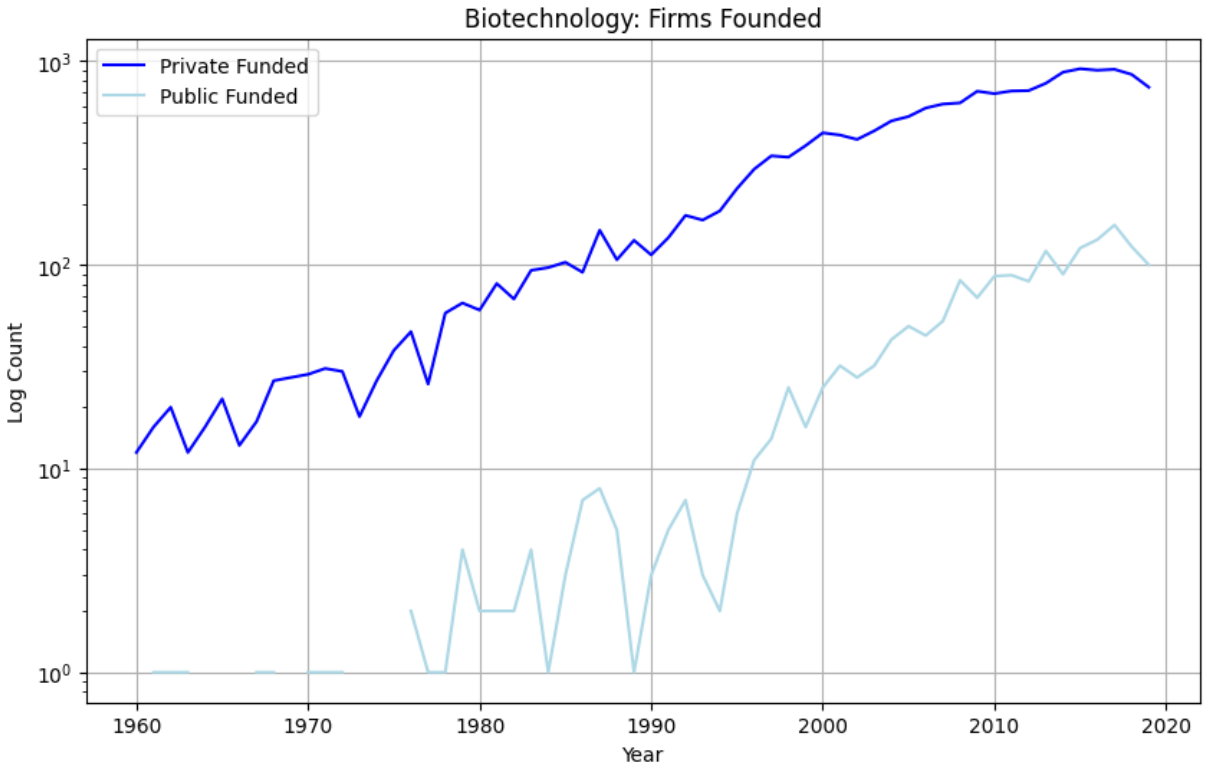

Artificial Intelligence (AI): Firms Founded
Count
Year
0
250
500
750
1000
1250
1500
1970
1980
1990
2000
2010
2020

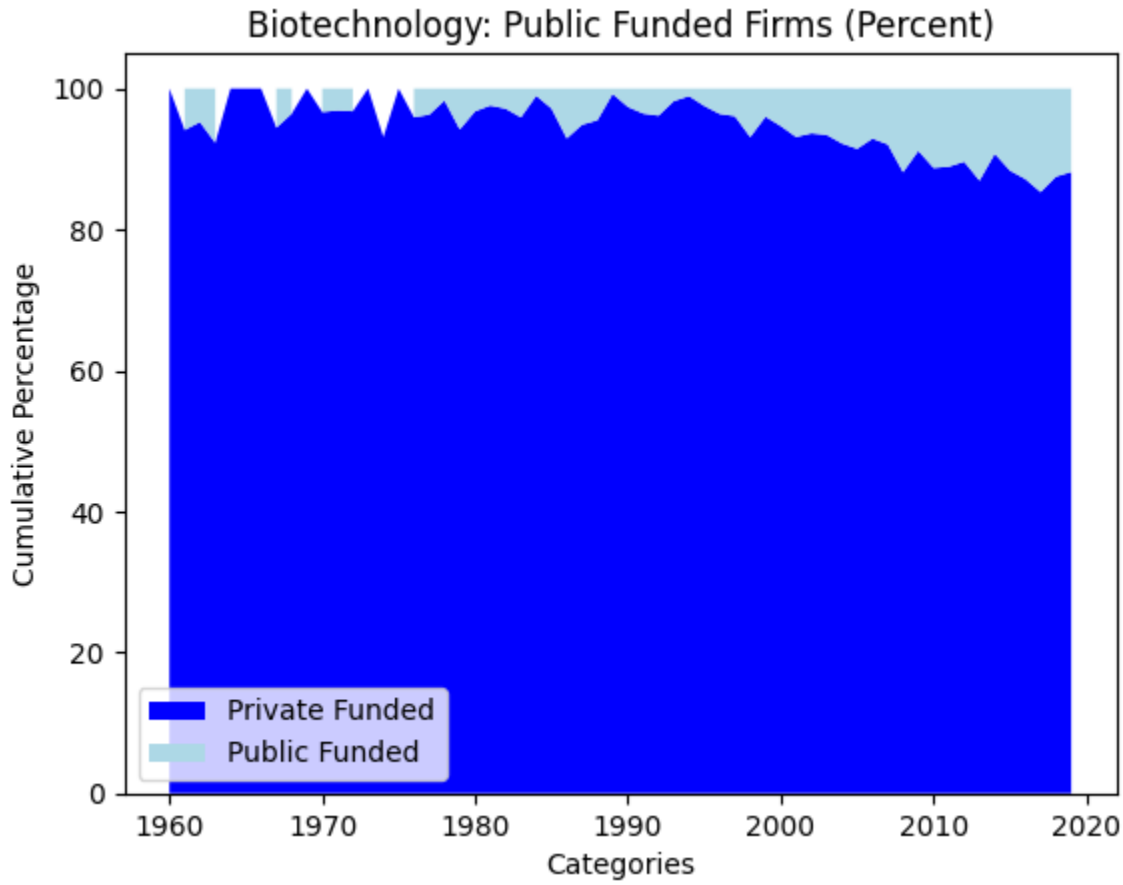

Artificial Intelligence (AI): Firms Founded
Private Funded
Public Funded
Log Count
Year
10^0
10^1
10^2
10^3
1970
1980
1990
2000
2010
2020



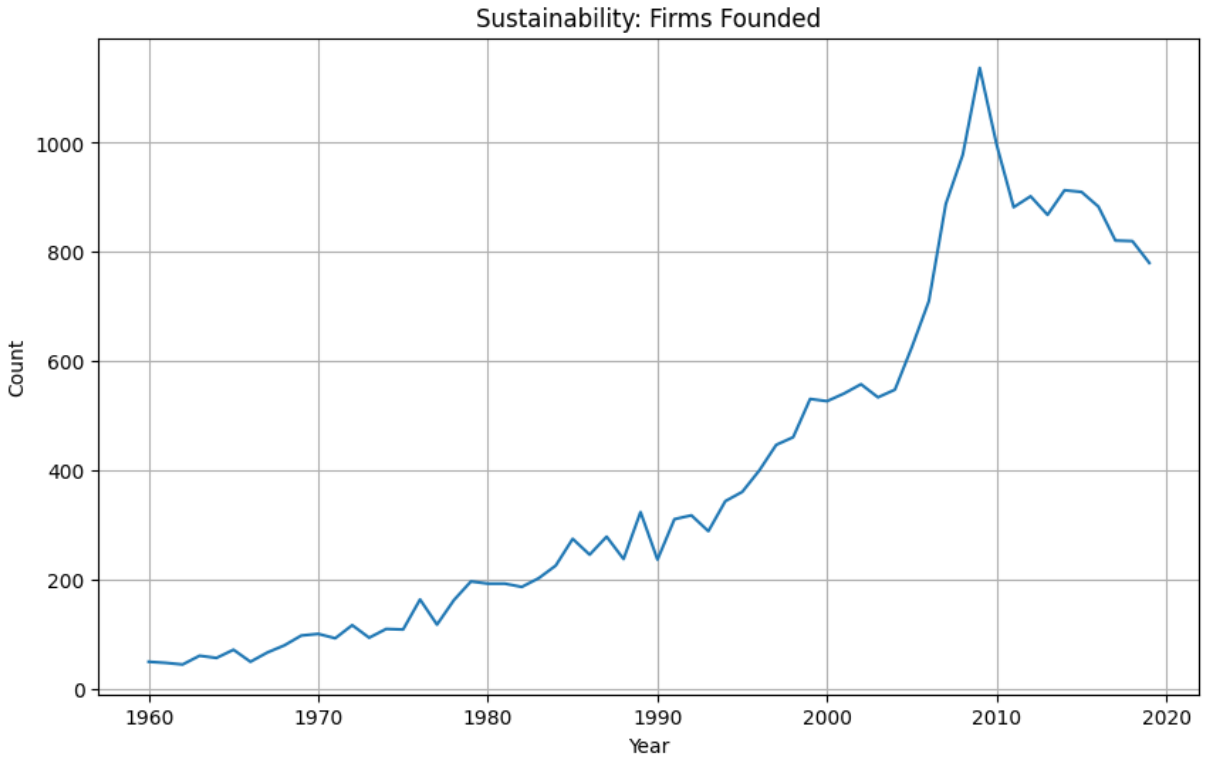
In [ ]:
```
visualise_industry_group(firms, 'Biotechnology', fed_rate)
```

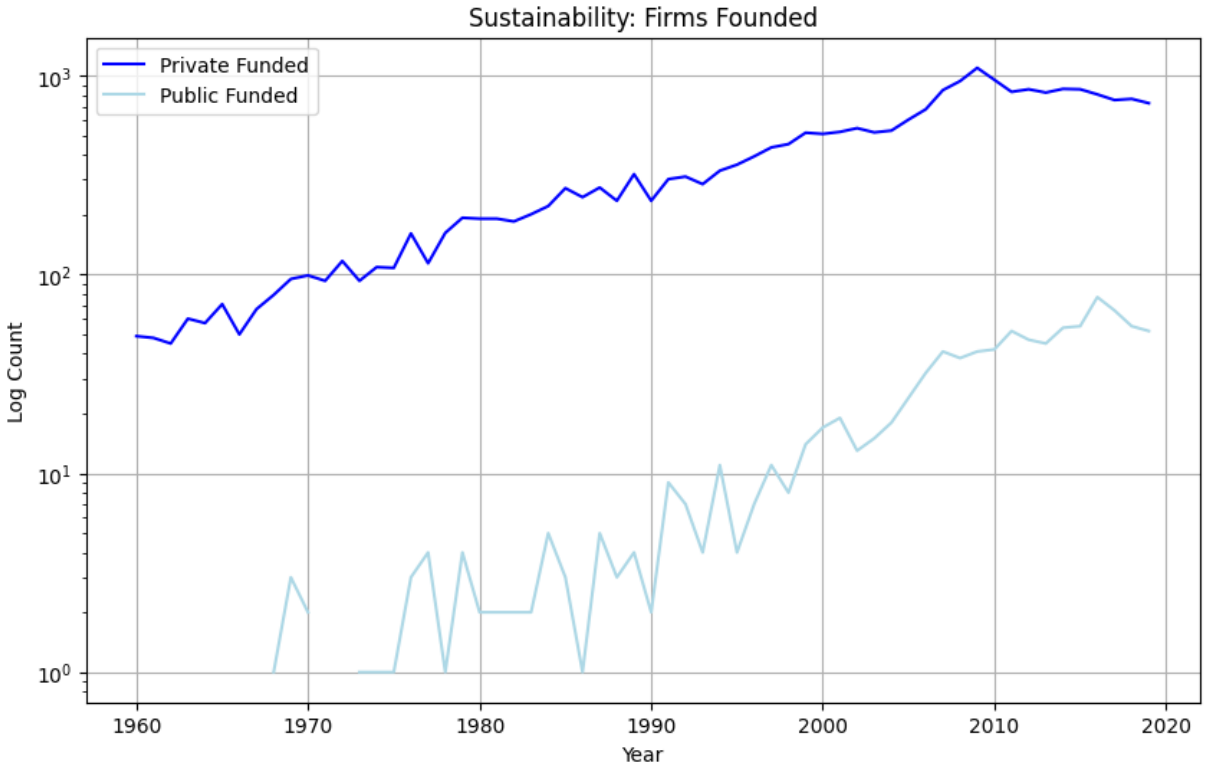


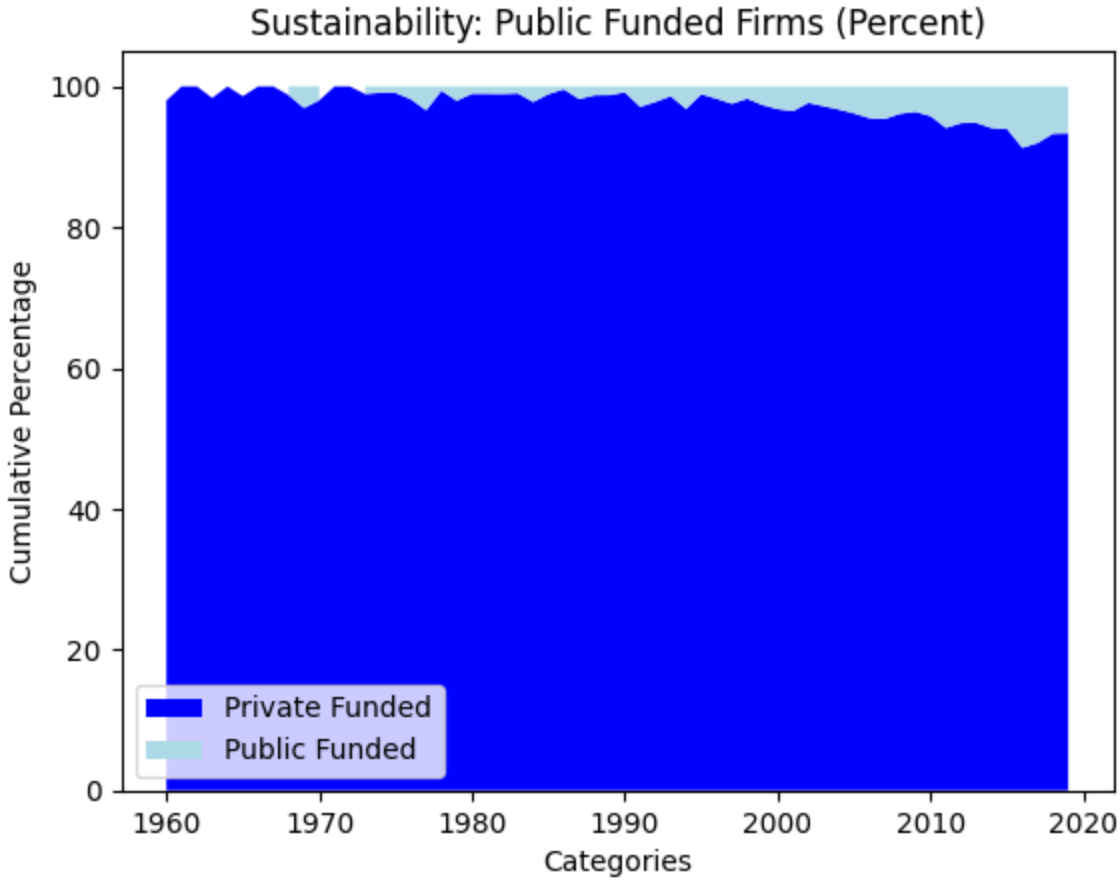


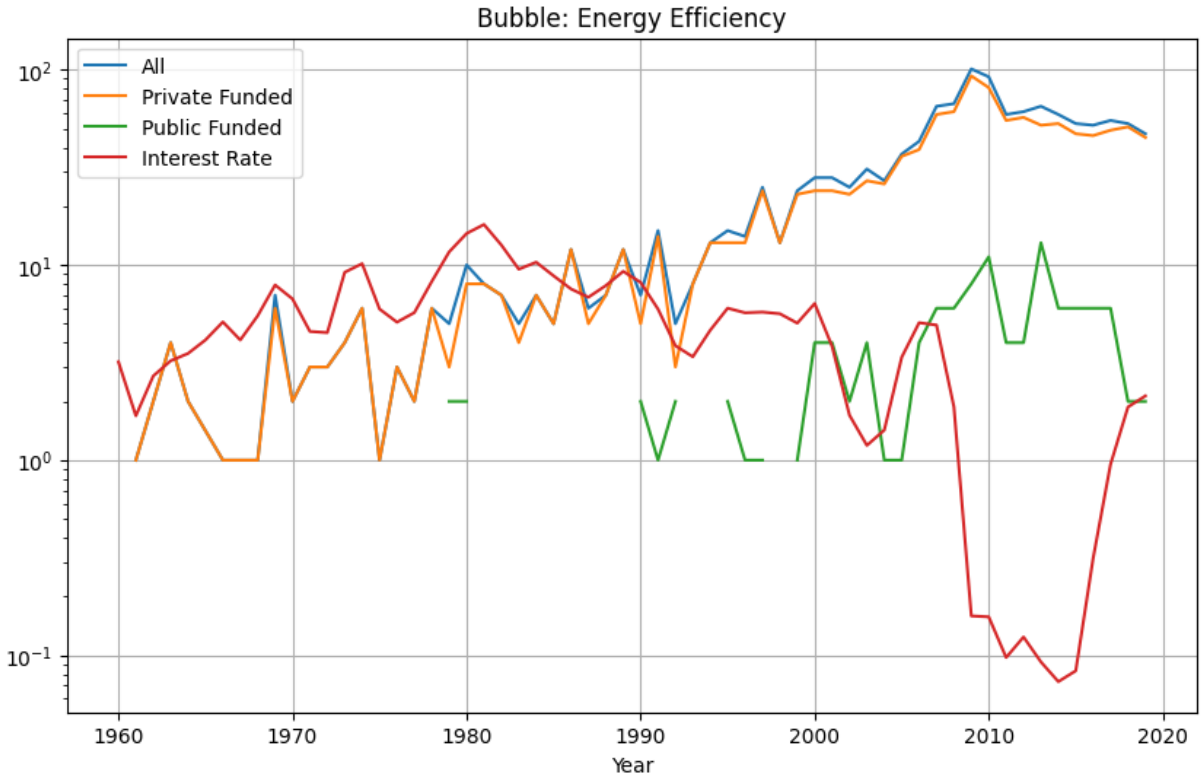


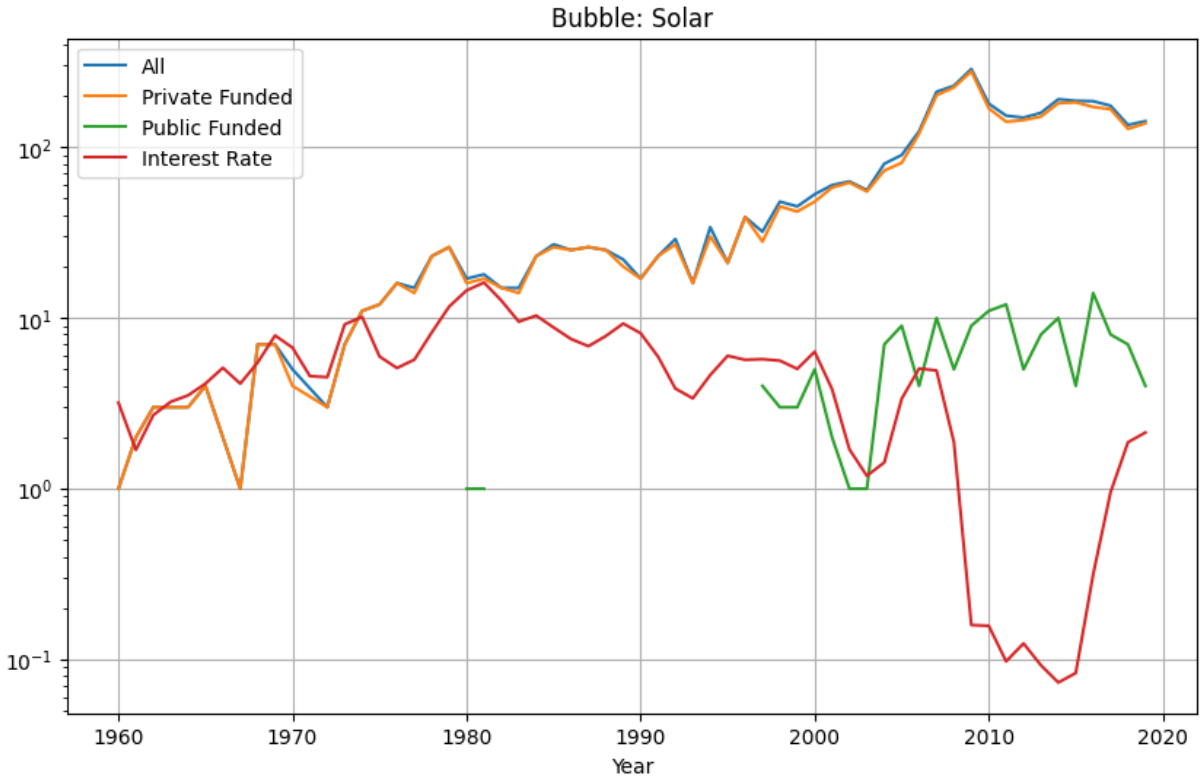
In [ ]:
```
visualise_industry_group(firms, 'Sustainability', fed_rate)
```


Sustainability: Firms Founded
Count
Year
0
200
400
600
800
1000
1960
1970
1980
1990
2000
2010
2020

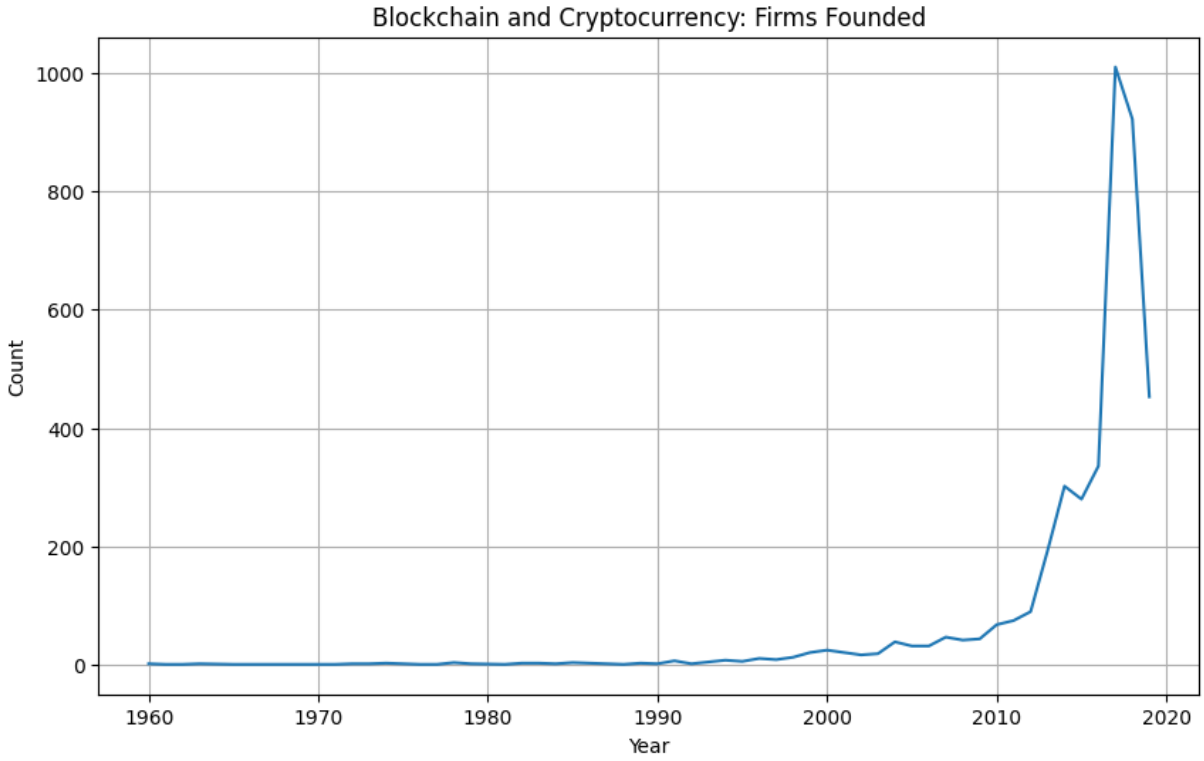

Sustainability: Firms Founded
Private Funded
Public Funded
Log Count
Year
10^0
10^1
10^2
10^3
1960
1970
1980
1990
2000
2010
2020

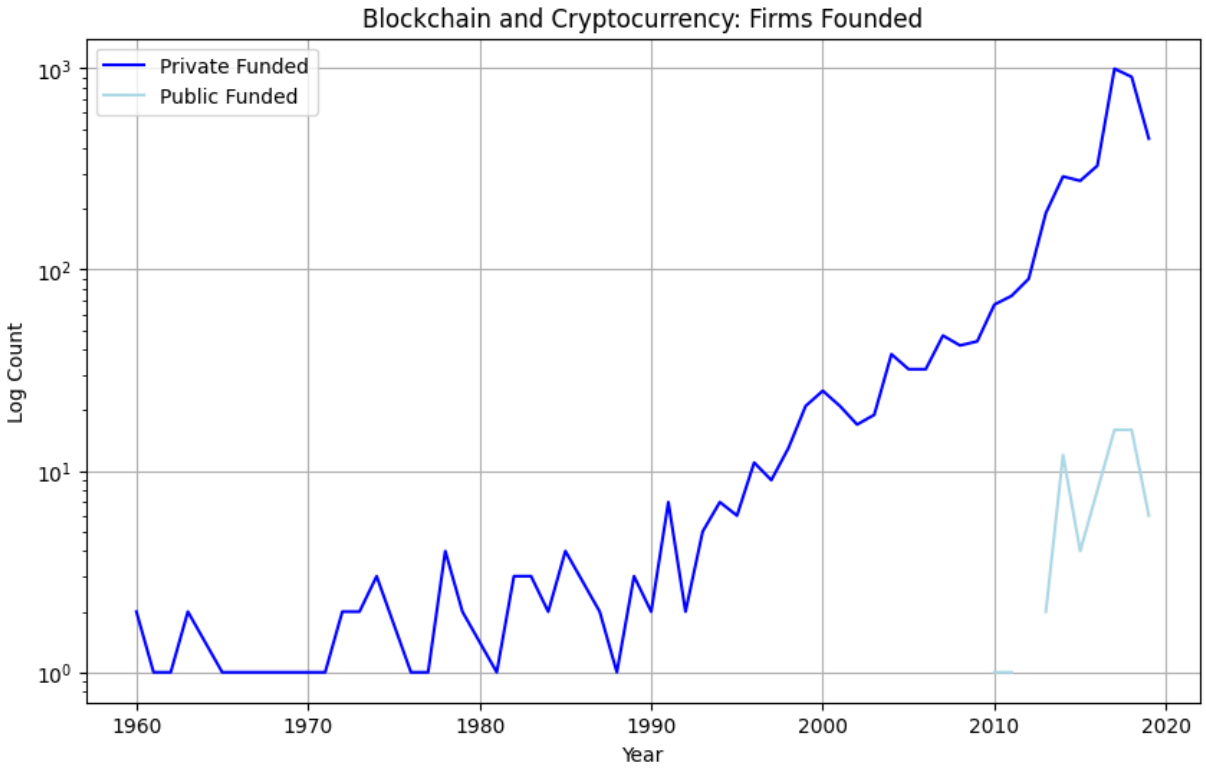

Sustainability: Public Funded Firms (Percent)
Cumulative Percentage
Categories
Private Funded
Public Funded
0
20
40
60
80
100
1960
1970
1980
1990
2000
2010
2020

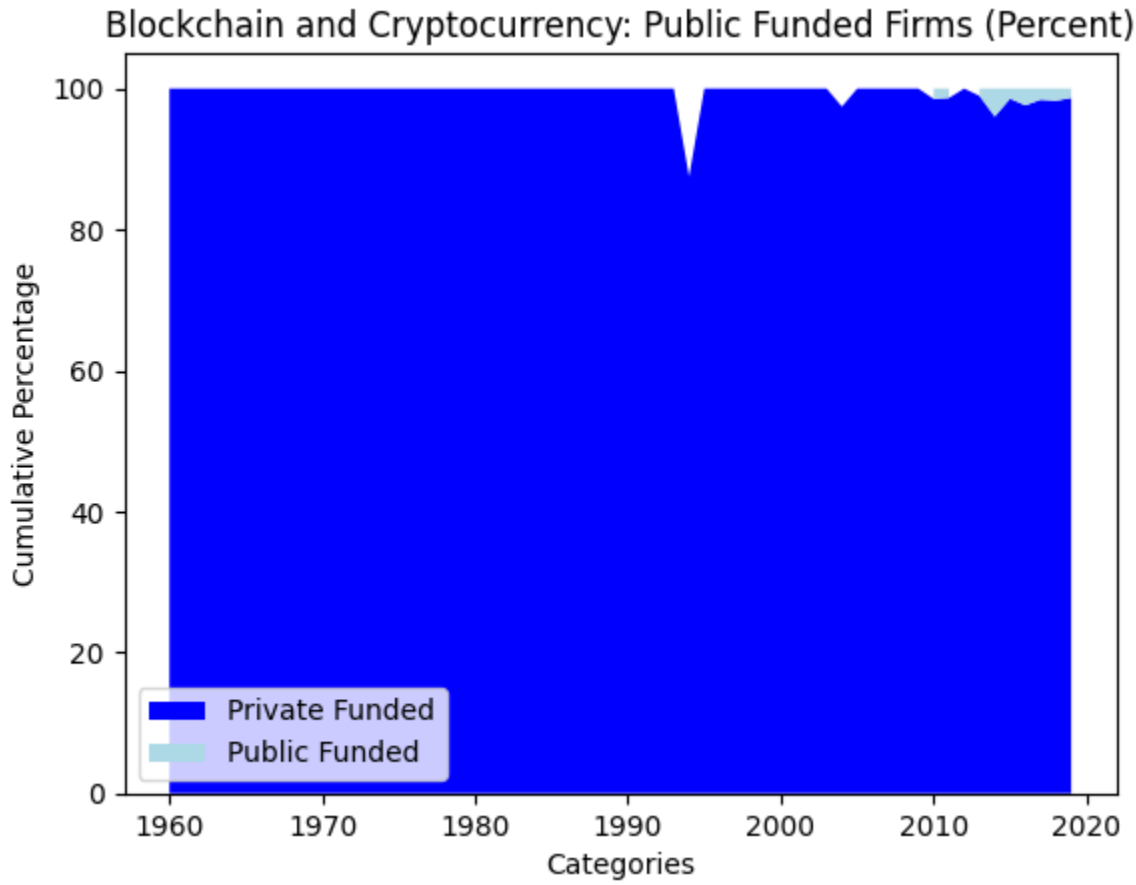

Bubble: Energy Efficiency
All
Private Funded
Public Funded
Interest Rate
Year
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020

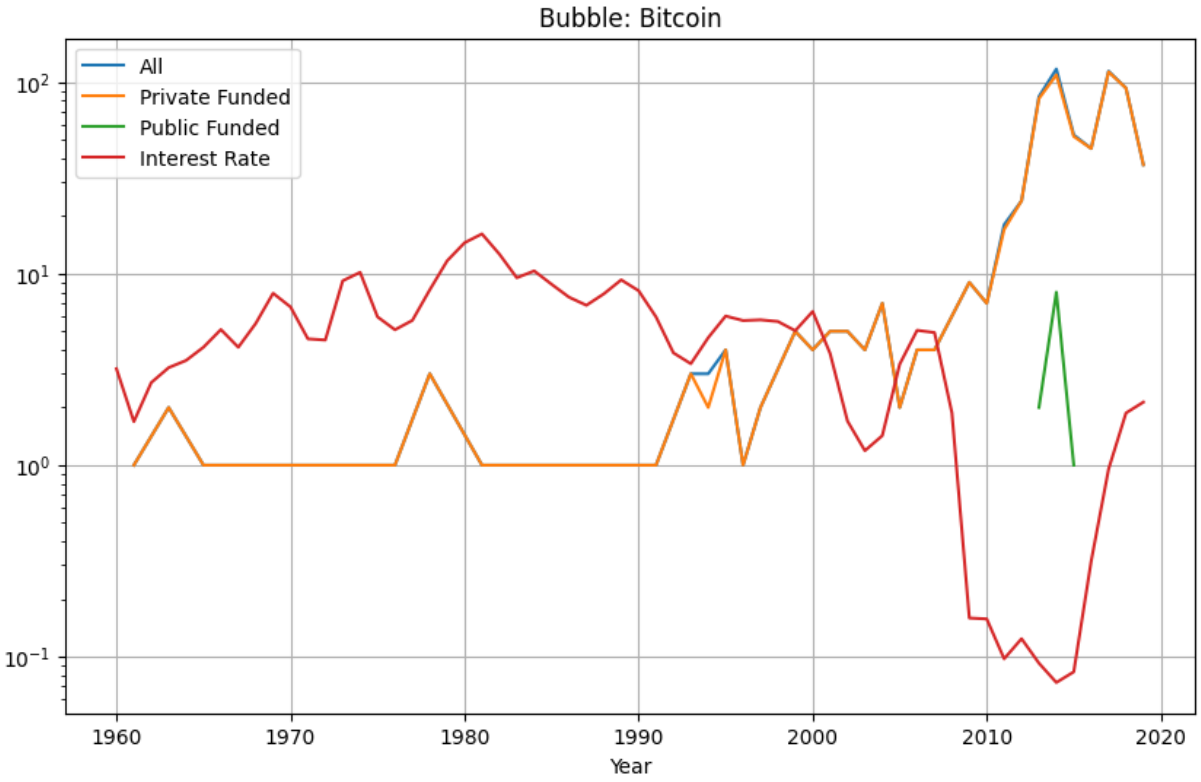


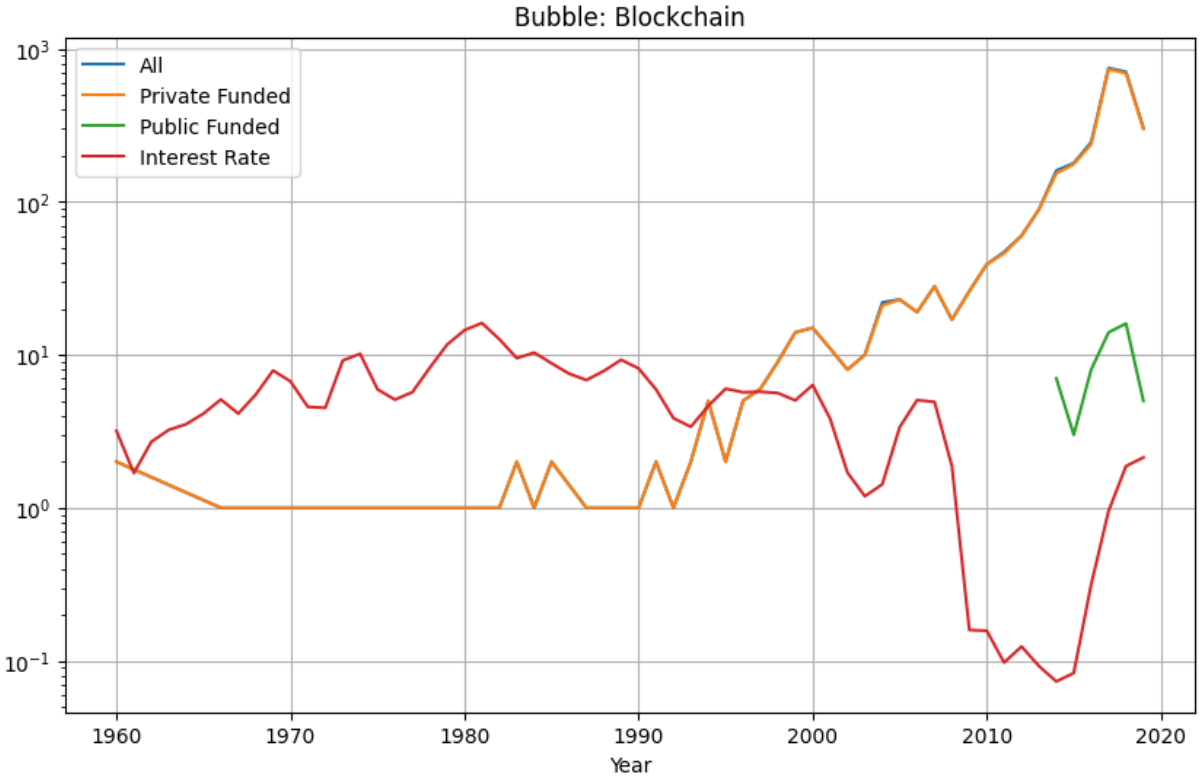
In [ ]:
```
visualise_industry_group(firms, 'Blockchain and Cryptocurrency', fed_rate)
```

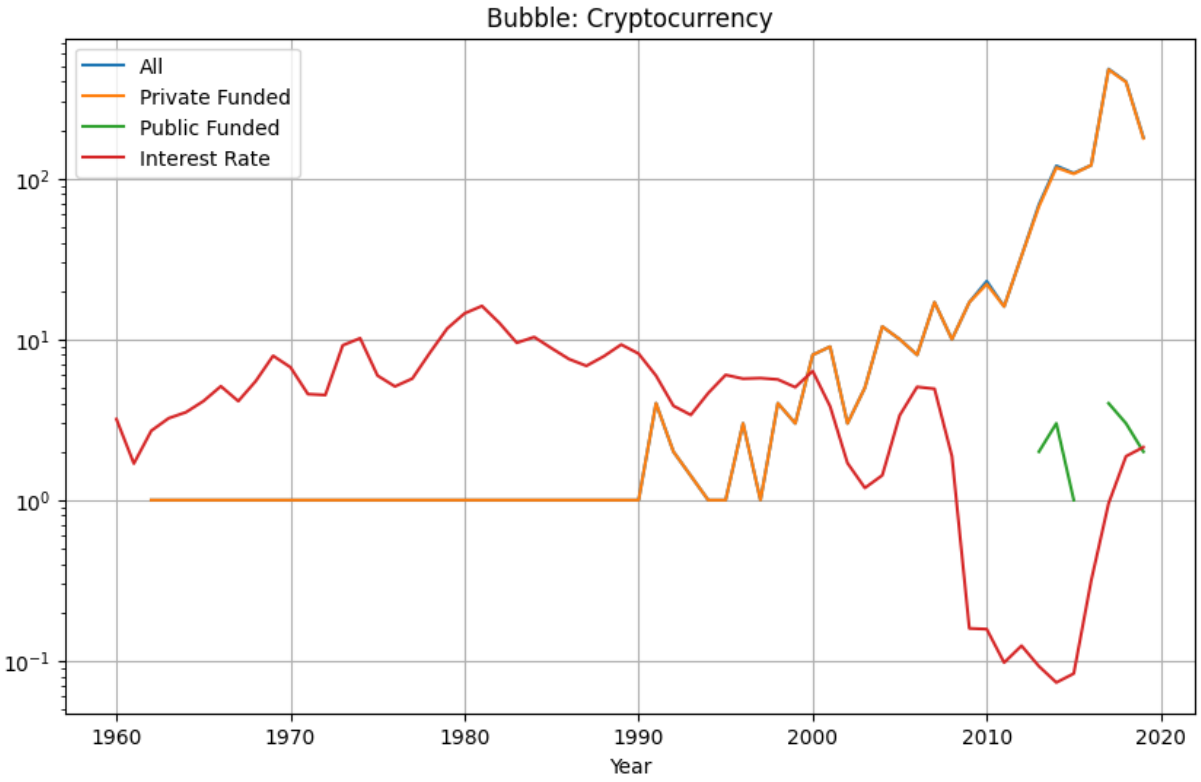



Blockchain and Cryptocurrency: Firms Founded
Private Funded
Public Funded
Log Count
Year
1960
1970
1980
1990
2000
2010
2020

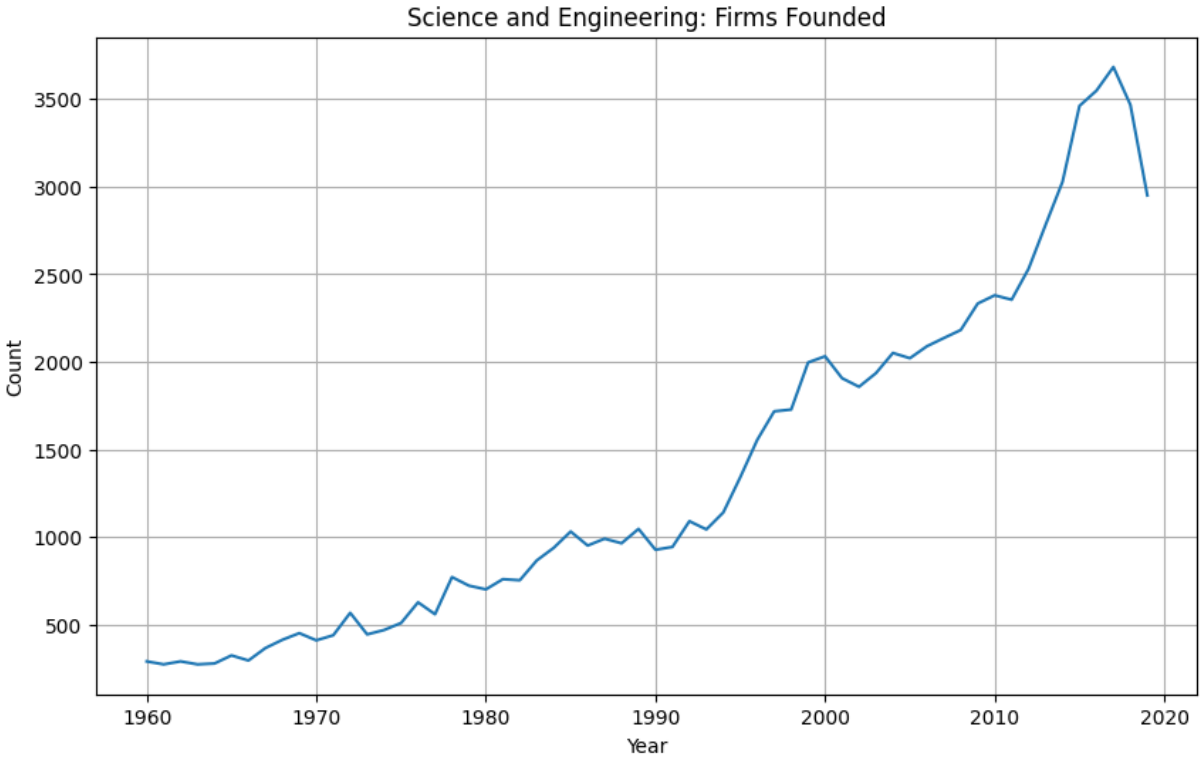

Blockchain and Cryptocurrency: Public Funded Firms (Percent)
Cumulative Percentage
Categories
Private Funded
Public Funded
0
20
40
60
80
100
1960
1970
1980
1990
2000
2010
2020

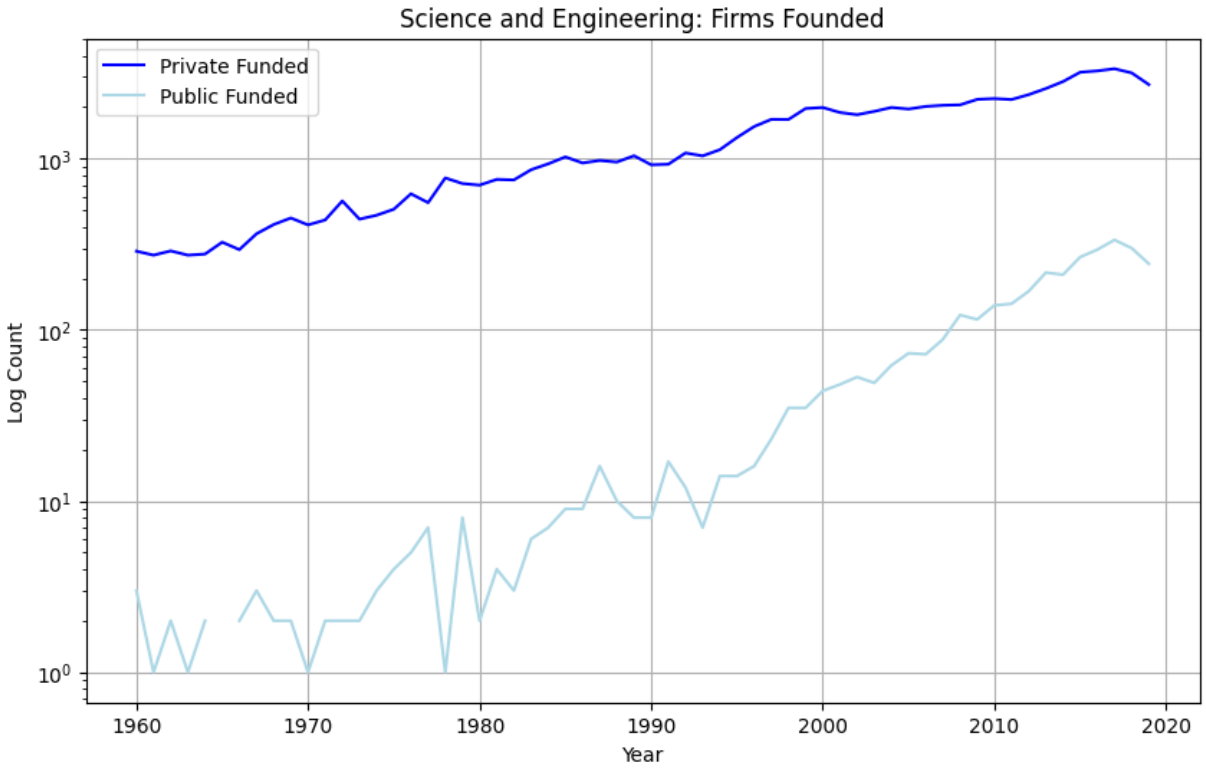

Bubble: Bitcoin
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

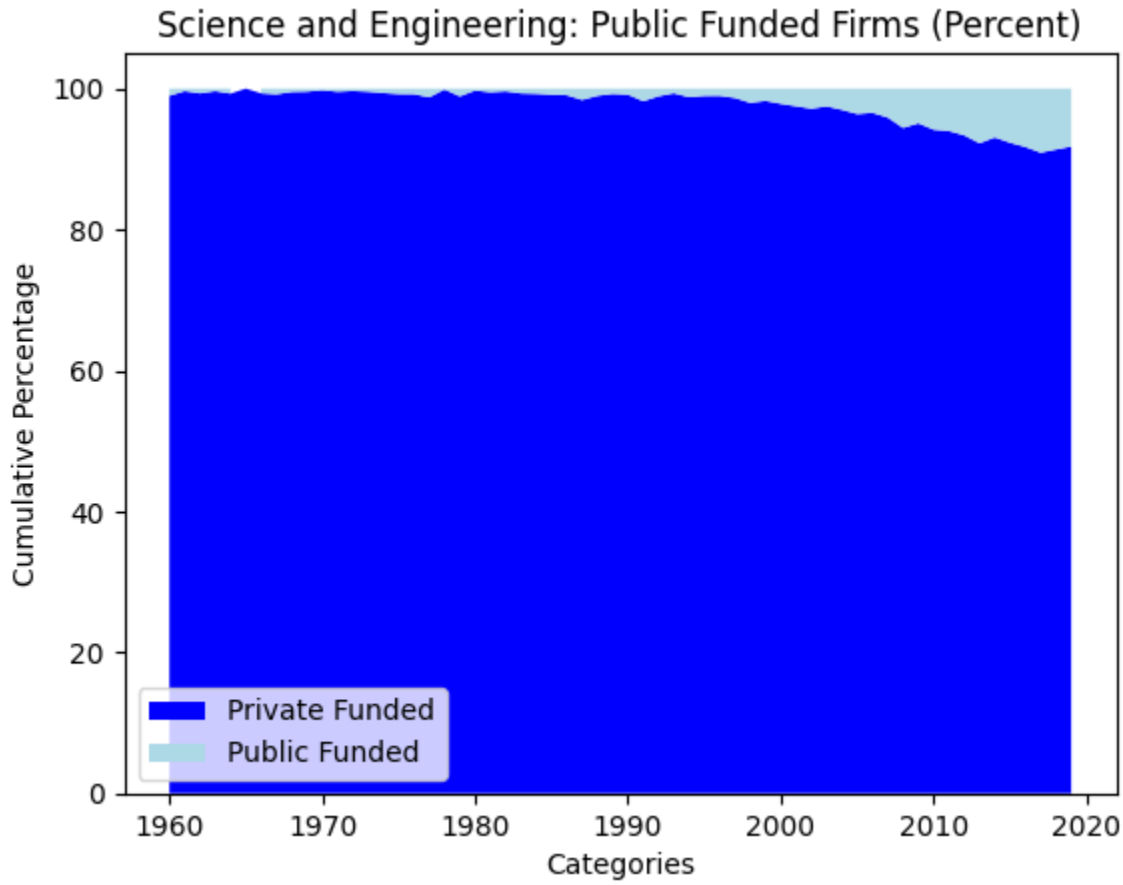

Bubble: Blockchain
All
Private Funded
Public Funded
Interest Rate
10^3
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

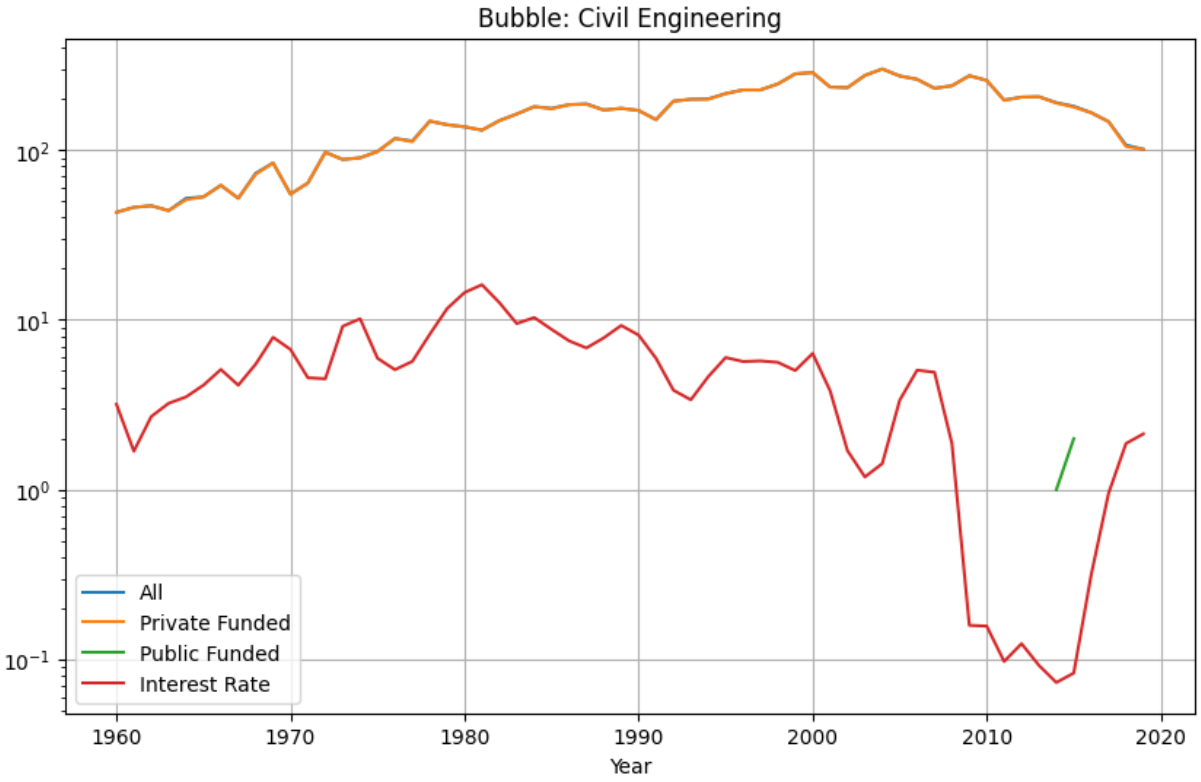


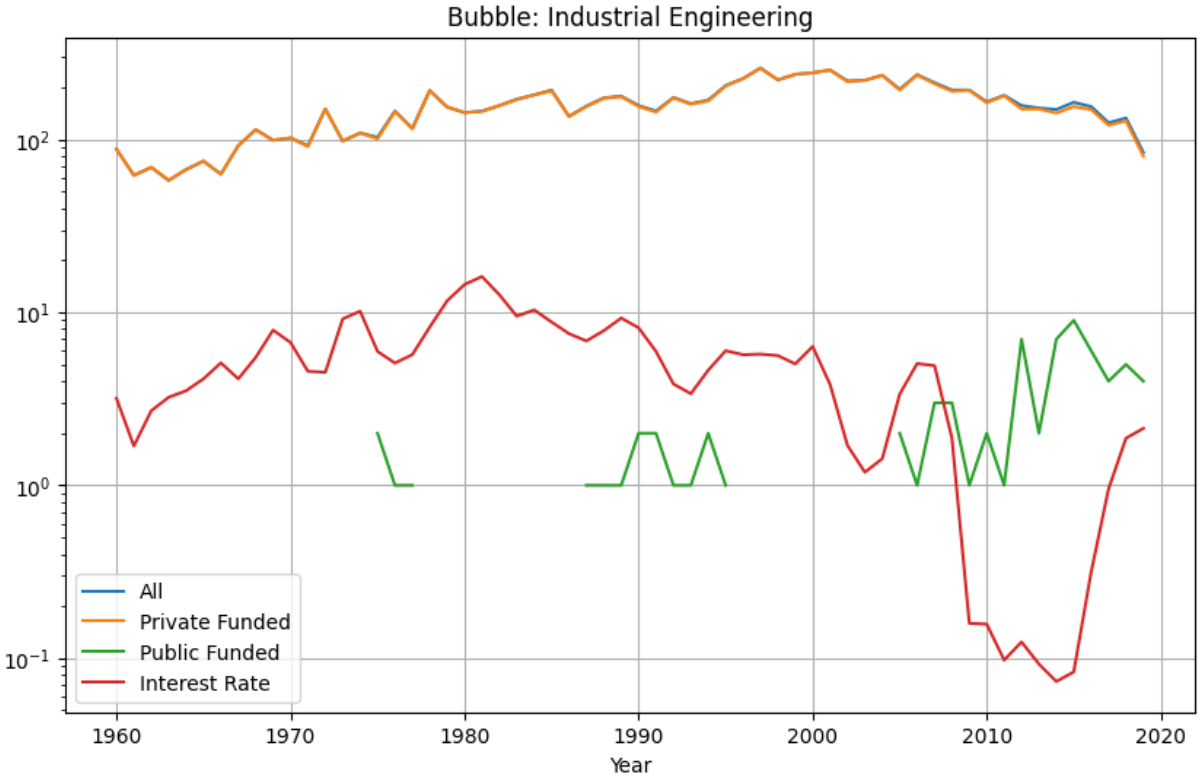
In [ ]:
```
visualise_industry_group(firms, 'Science and Engineering', fed_rate)
```

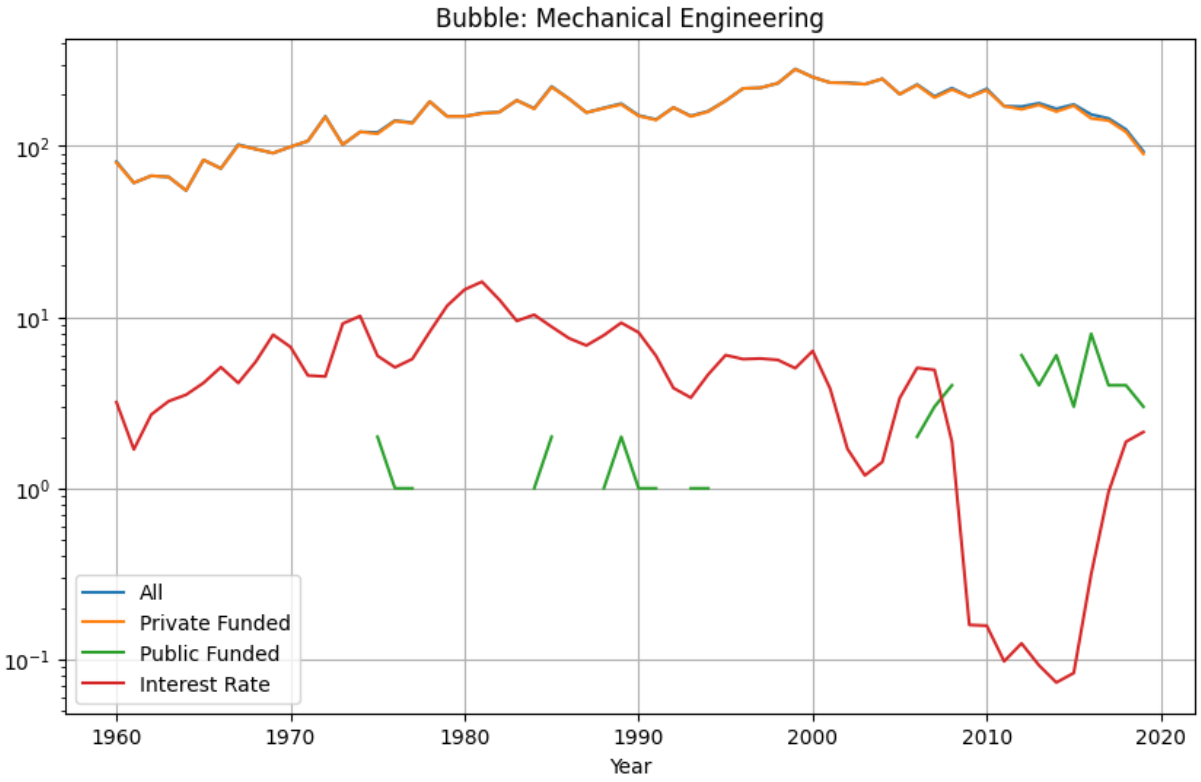


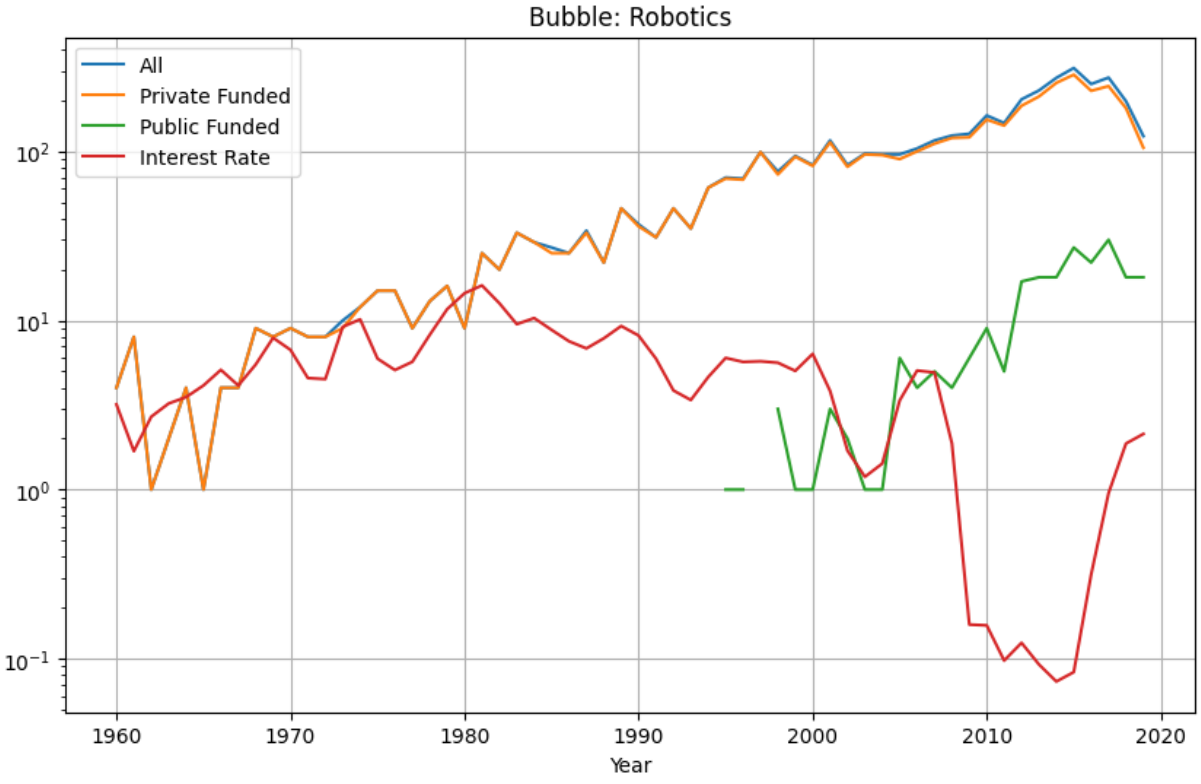

Science and Engineering: Firms Founded
Private Funded
Public Funded
Log Count
10^3
10^2
10^1
10^0
1960
1970
1980
1990
2000
2010
2020
Year

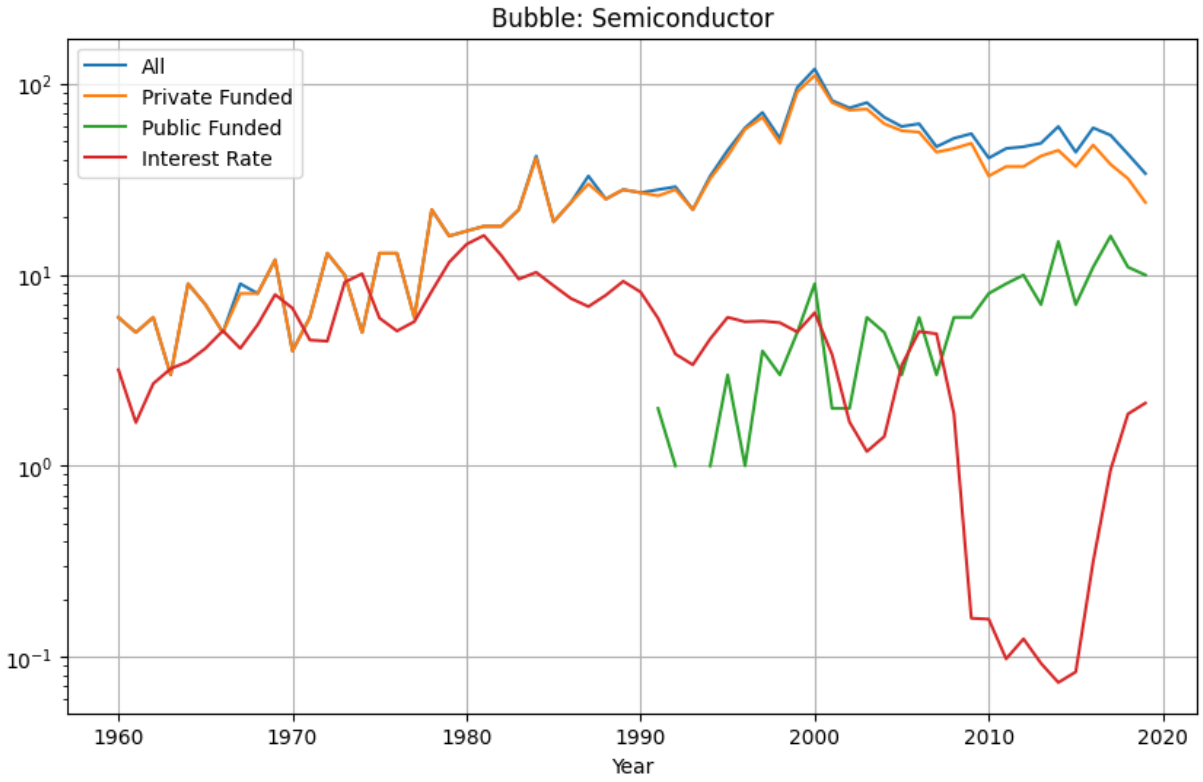

Science and Engineering: Public Funded Firms (Percent)
Cumulative Percentage
100
80
60
40
20
0
Private Funded
Public Funded
1960
1970
1980
1990
2000
2010
2020
Categories


Bubble: Civil Engineering
10^2
10^1
10^0
10^{-1}
All
Private Funded
Public Funded
Interest Rate
1960
1970
1980
1990
2000
2010
2020
Year

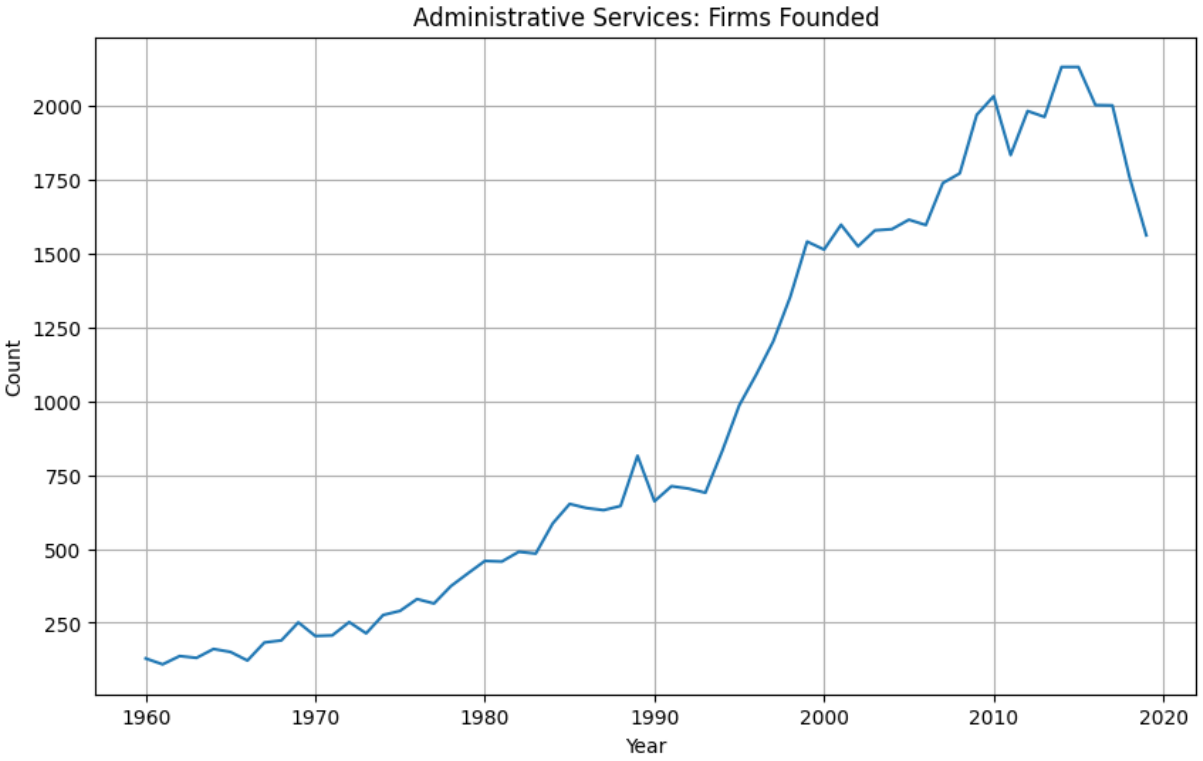

Bubble: Industrial Engineering
10^2
10^1
10^0
10^{-1}
All
Private Funded
Public Funded
Interest Rate
1960
1970
1980
1990
2000
2010
2020
Year

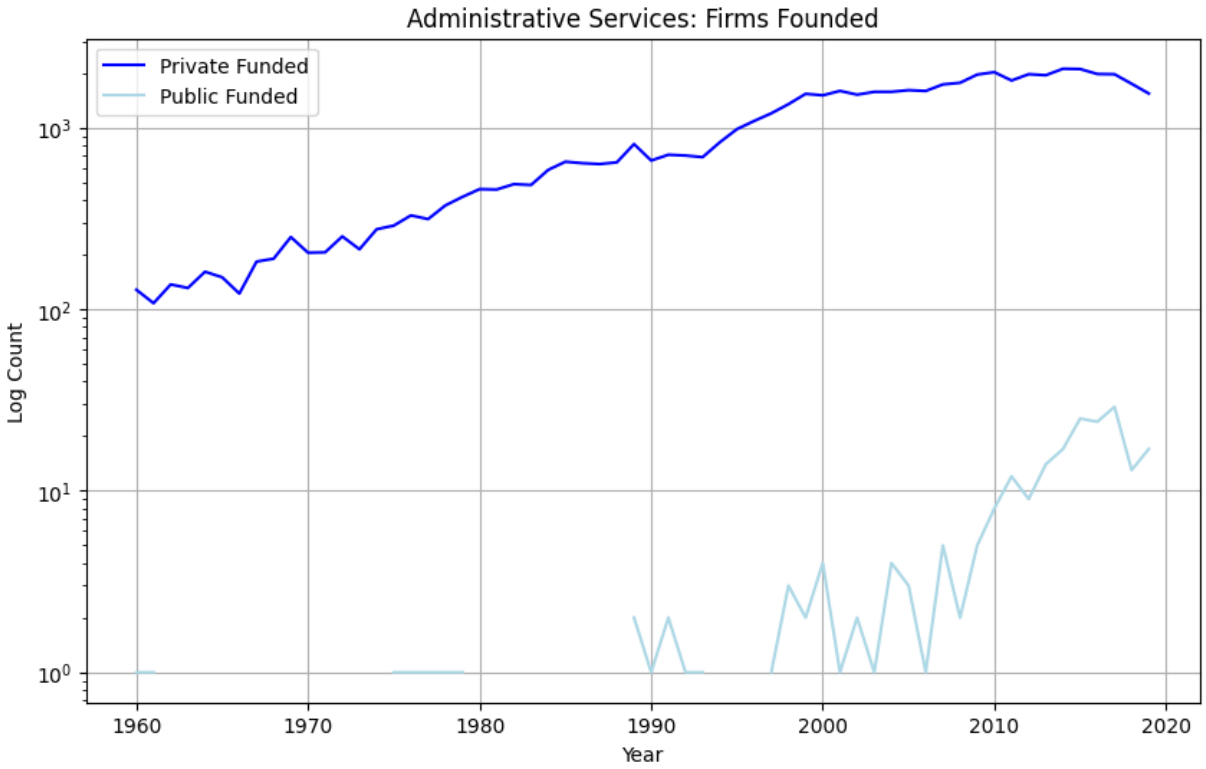

Bubble: Mechanical Engineering
All
Private Funded
Public Funded
Interest Rate
Year

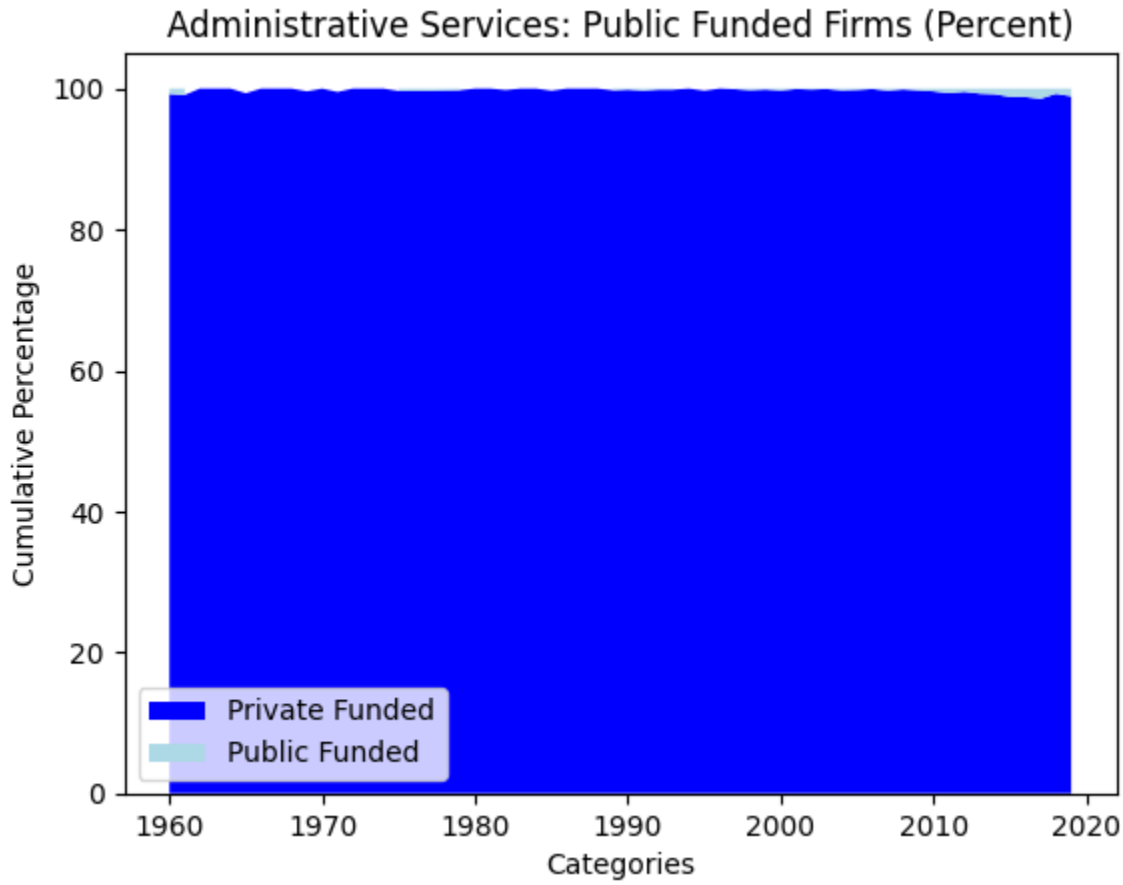

Bubble: Robotics
All
Private Funded
Public Funded
Interest Rate
Year

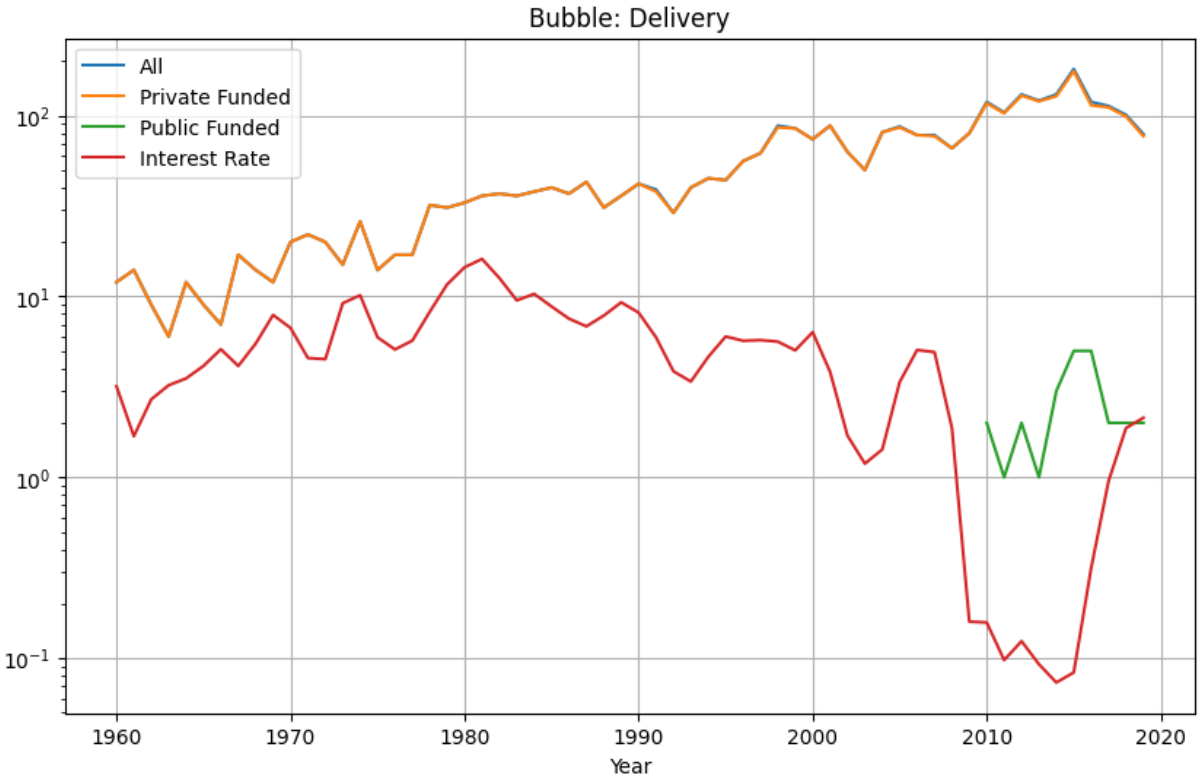


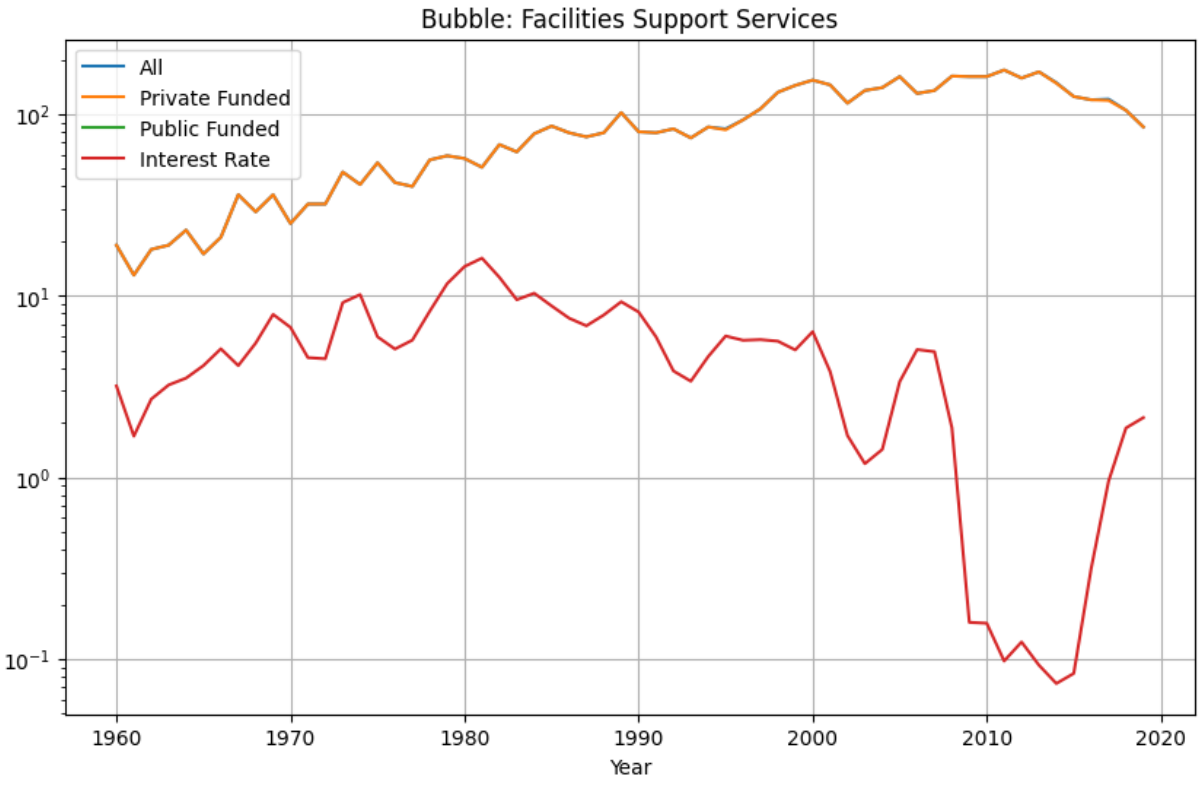
In [ ]:
```
visualise_industry_group(firms, 'Administrative Services', fed_rate)
```



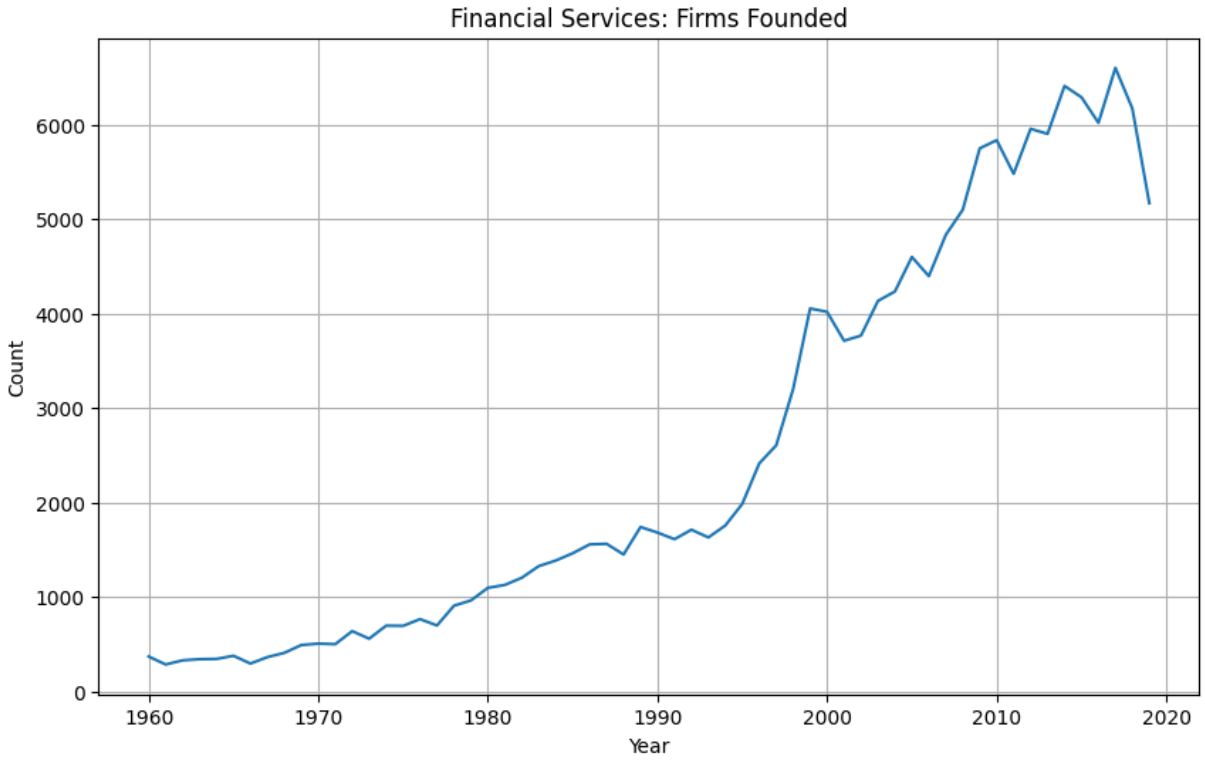

Administrative Services: Firms Founded
Private Funded
Public Funded
Log Count
10^3
10^2
10^1
10^0
1960
1970
1980
1990
2000
2010
2020
Year

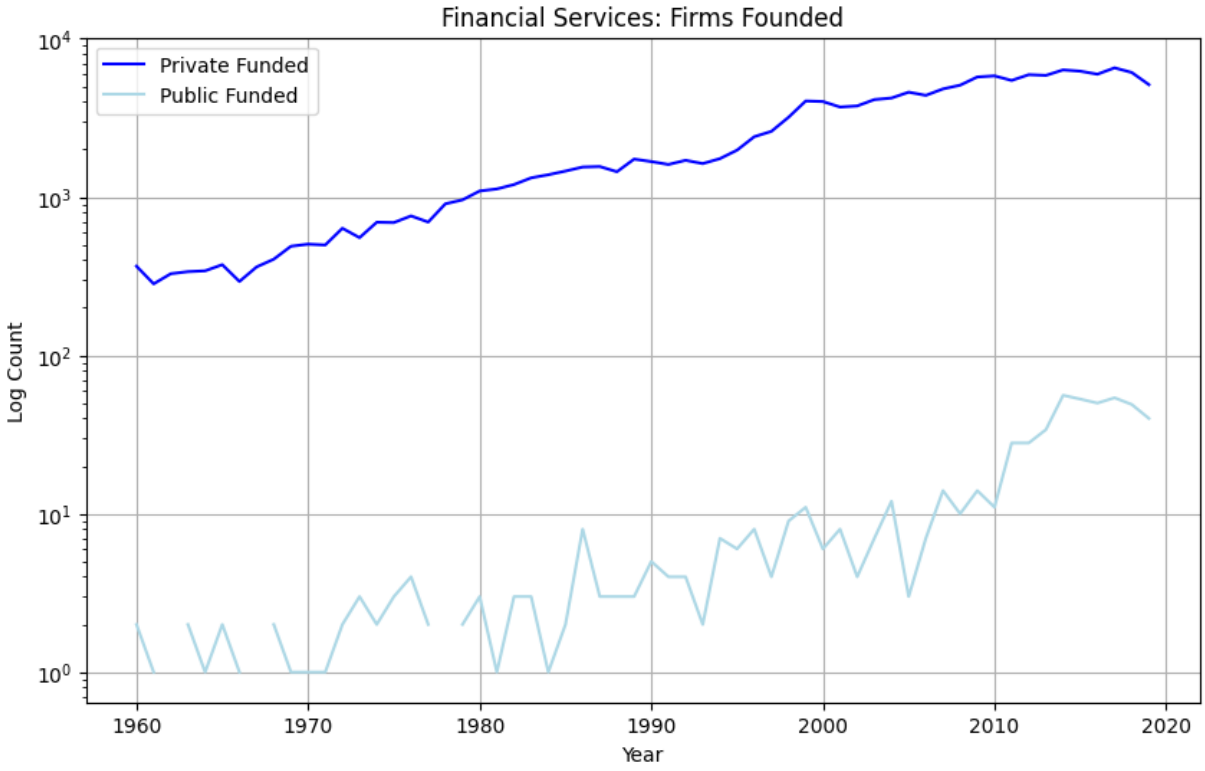

Administrative Services: Public Funded Firms (Percent)
Cumulative Percentage
100
80
60
40
20
0
Private Funded
Public Funded
1960
1970
1980
1990
2000
2010
2020
Categories

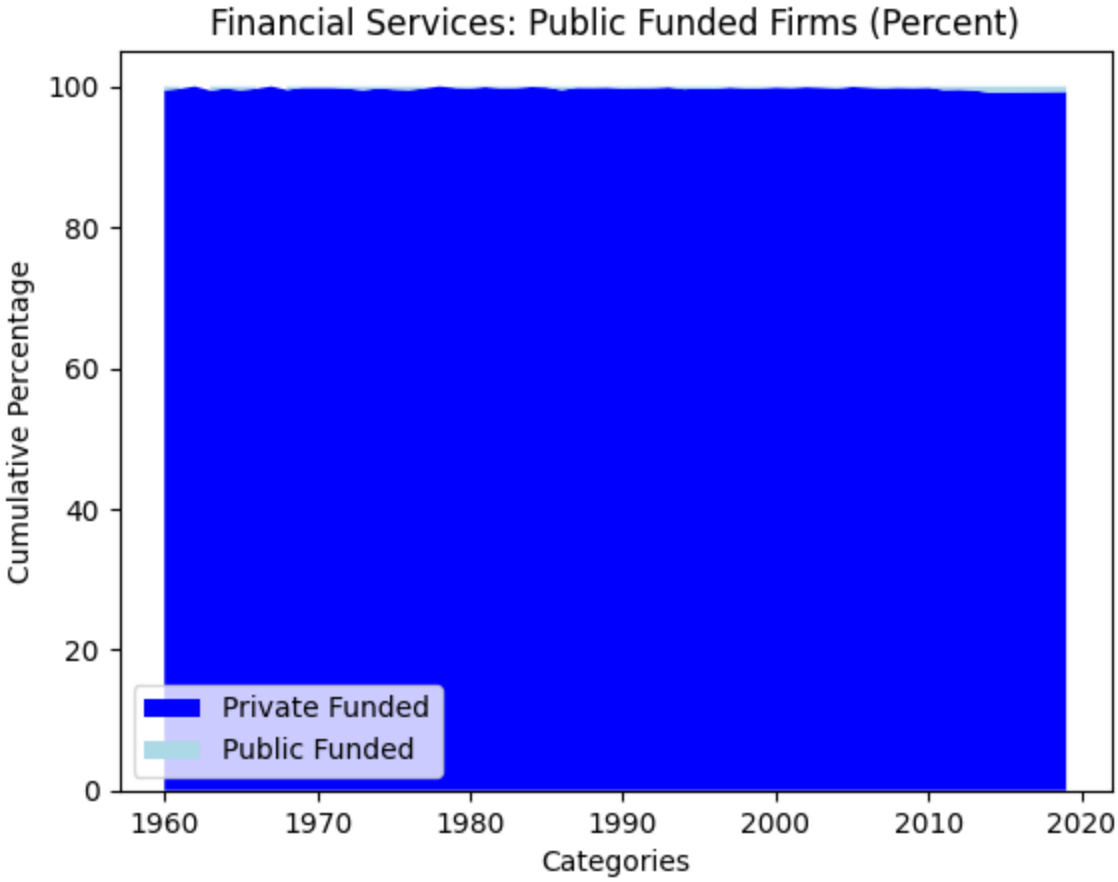


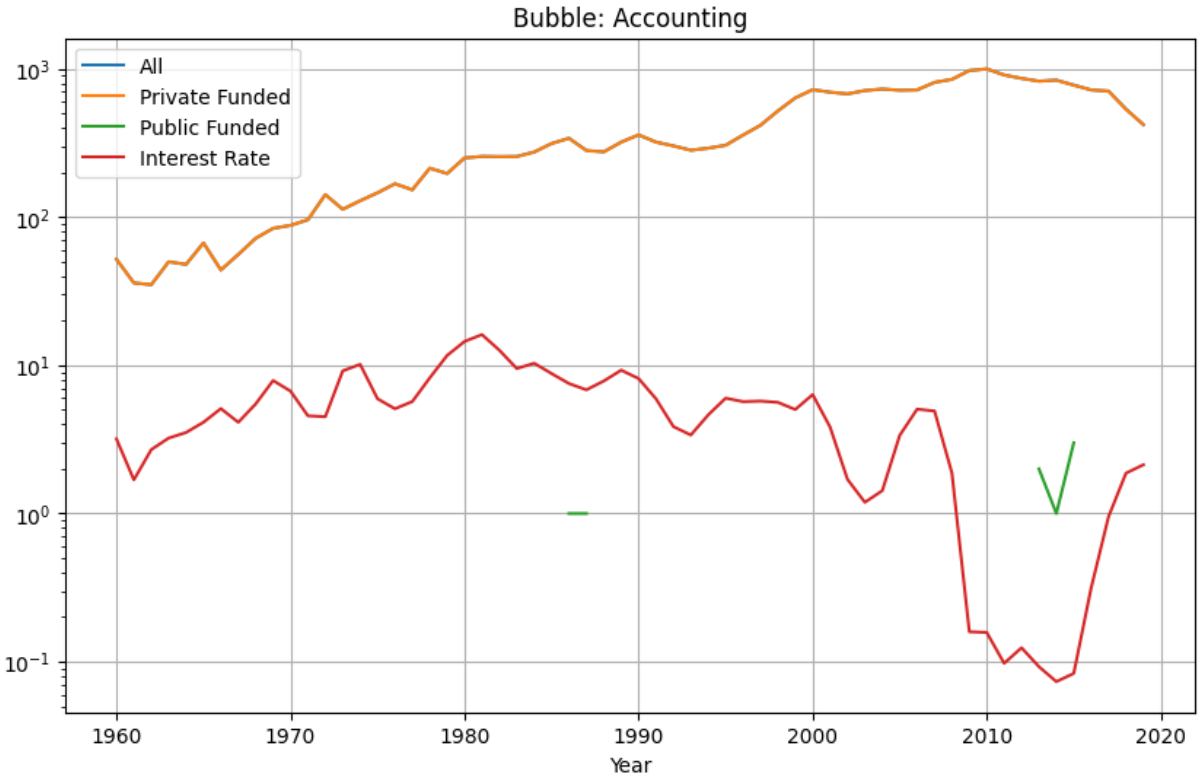


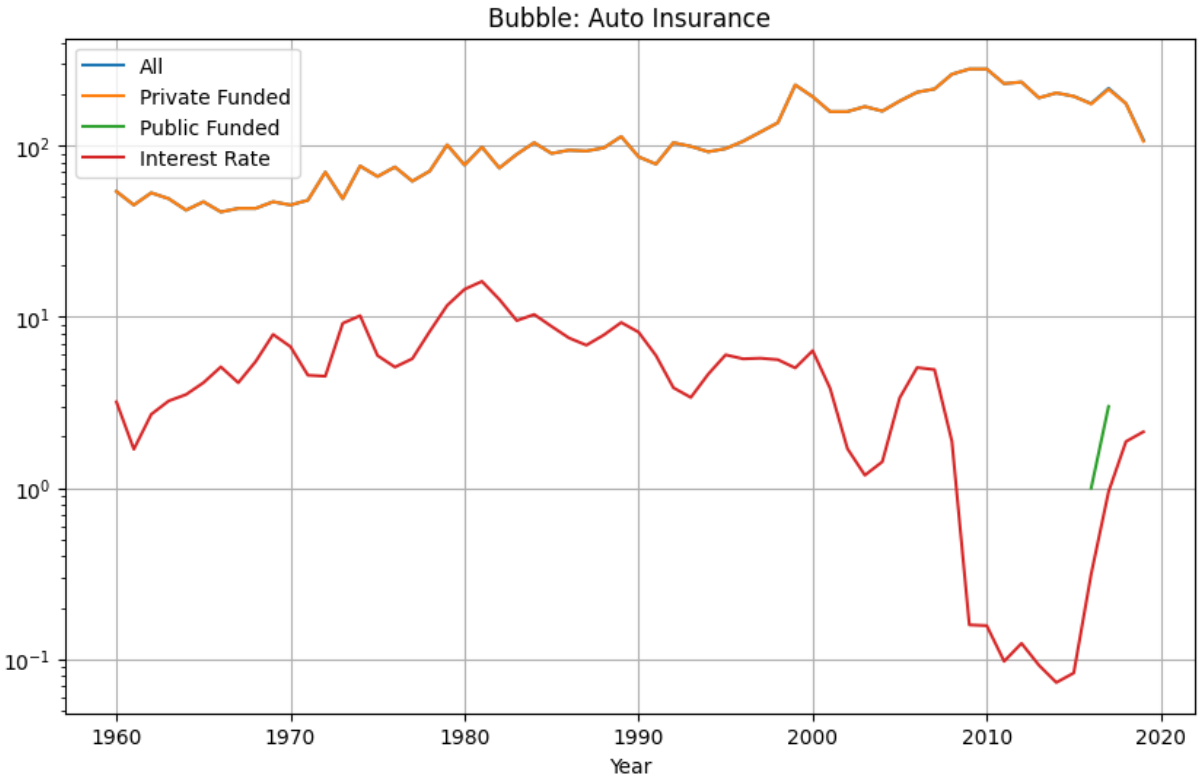
In [ ]:
```
visualise_industry_group(firms, 'Financial Services', fed_rate)
```

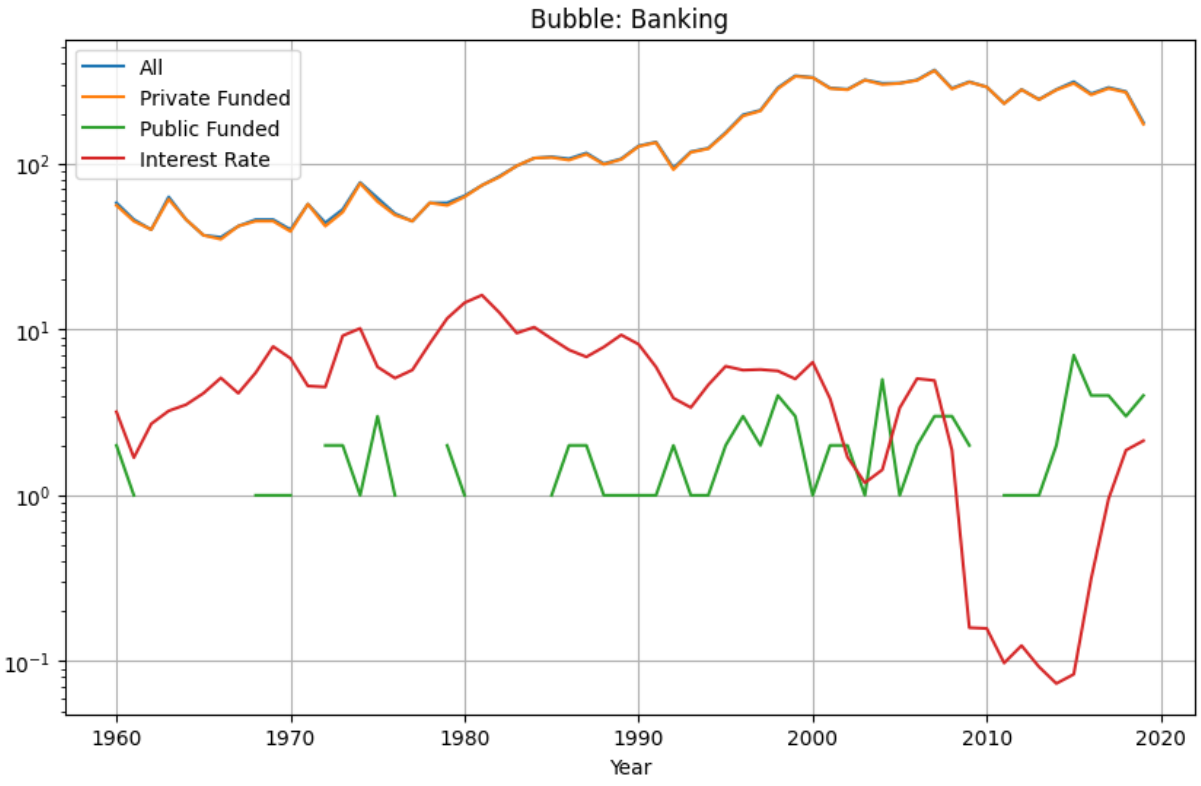

Financial Services: Firms Founded
Count
Year
0
1000
2000
3000
4000
5000
6000
1960
1970
1980
1990
2000
2010
2020

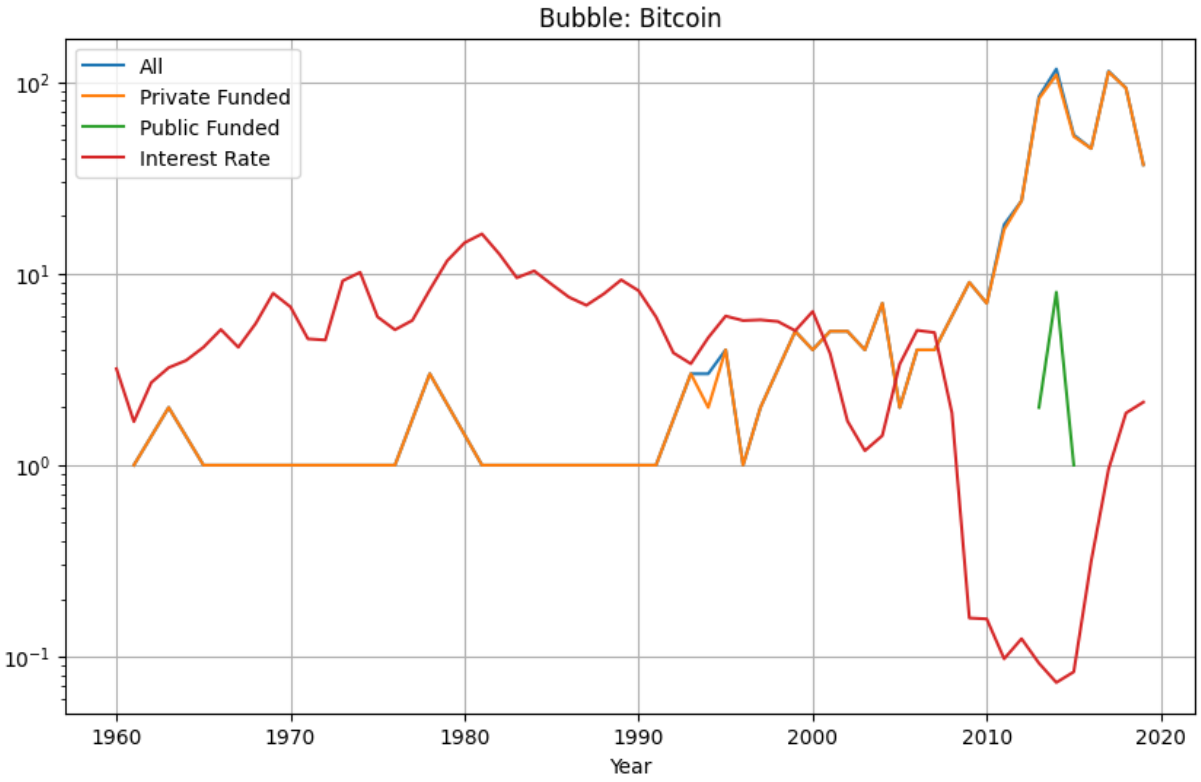

Financial Services: Firms Founded
Private Funded
Public Funded
Log Count
Year
10^0
10^1
10^2
10^3
10^4
1960
1970
1980
1990
2000
2010
2020

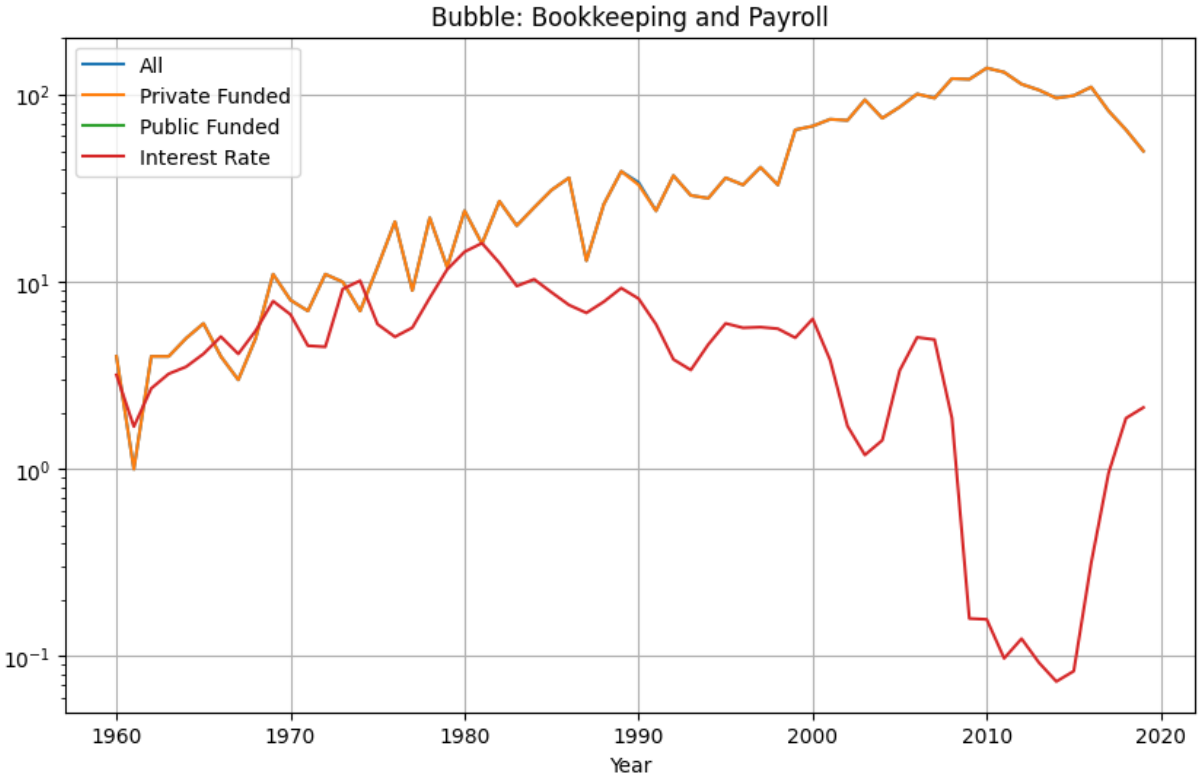

Financial Services: Public Funded Firms (Percent)
Cumulative Percentage
Categories
Private Funded
Public Funded
0
20
40
60
80
100
1960
1970
1980
1990
2000
2010
2020

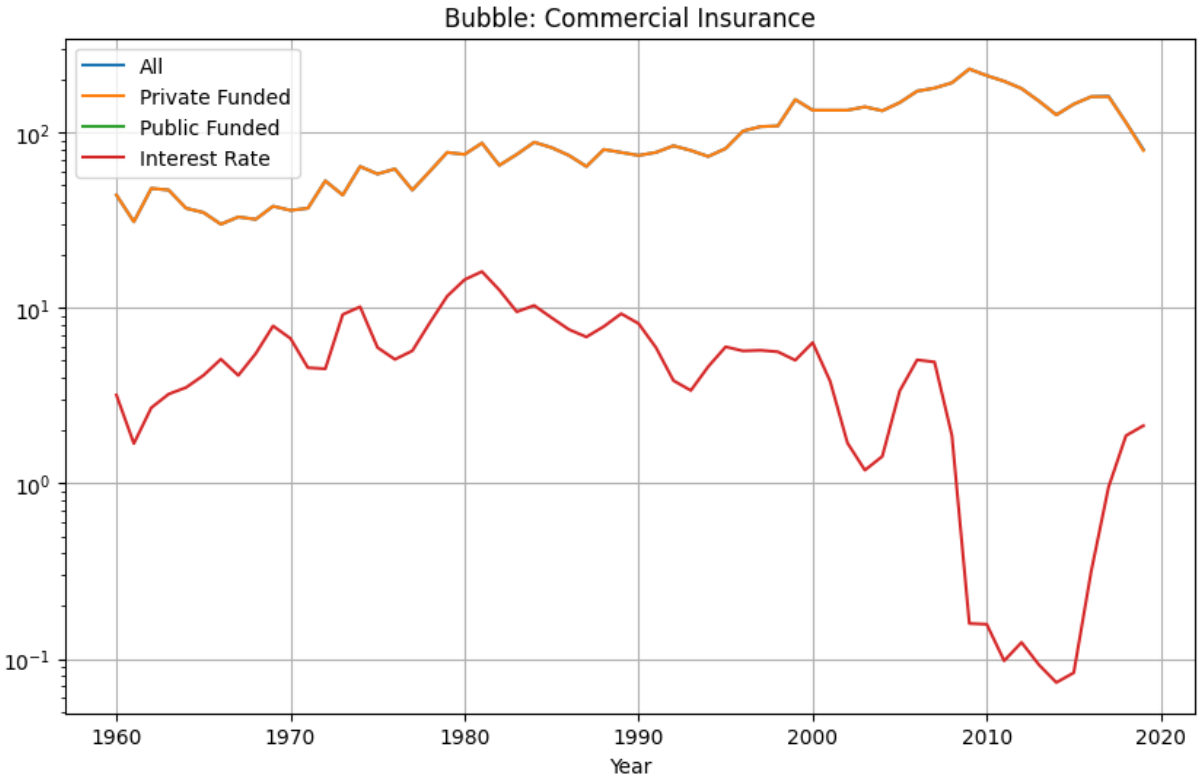

Bubble: Accounting
All
Private Funded
Public Funded
Interest Rate
Year
1960
1970
1980
1990
2000
2010
2020

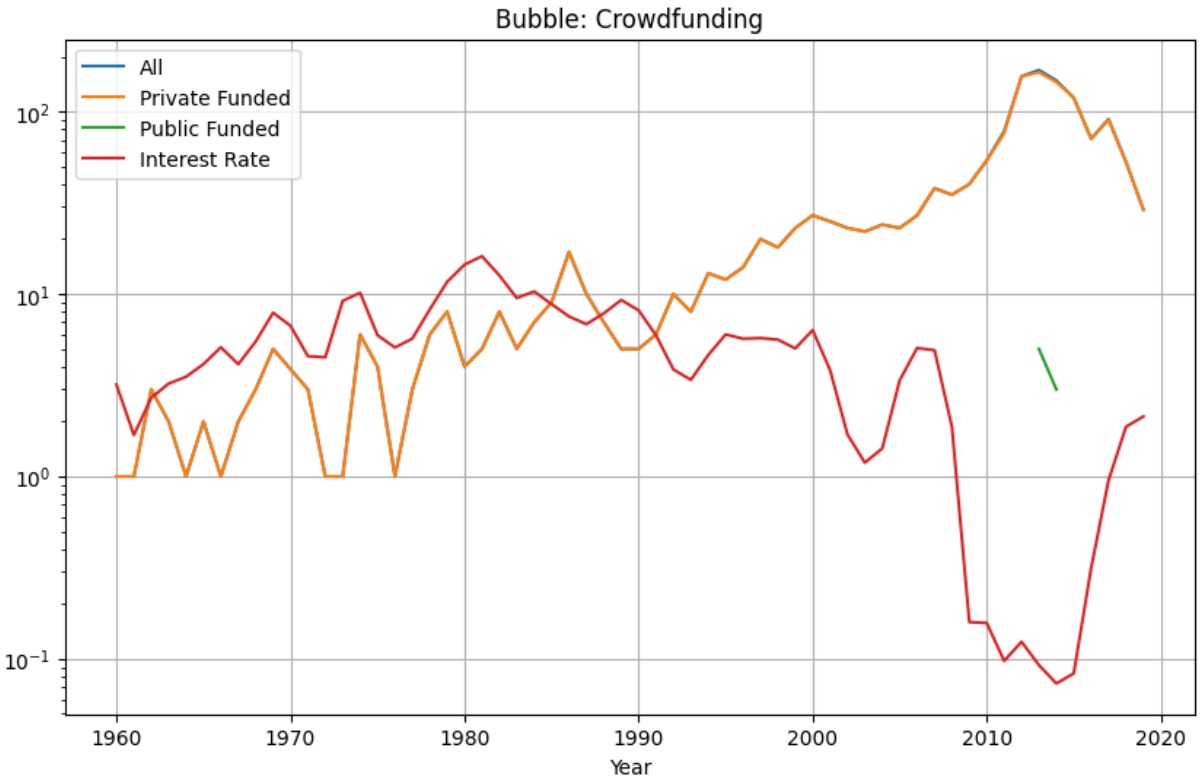

Bubble: Auto Insurance
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

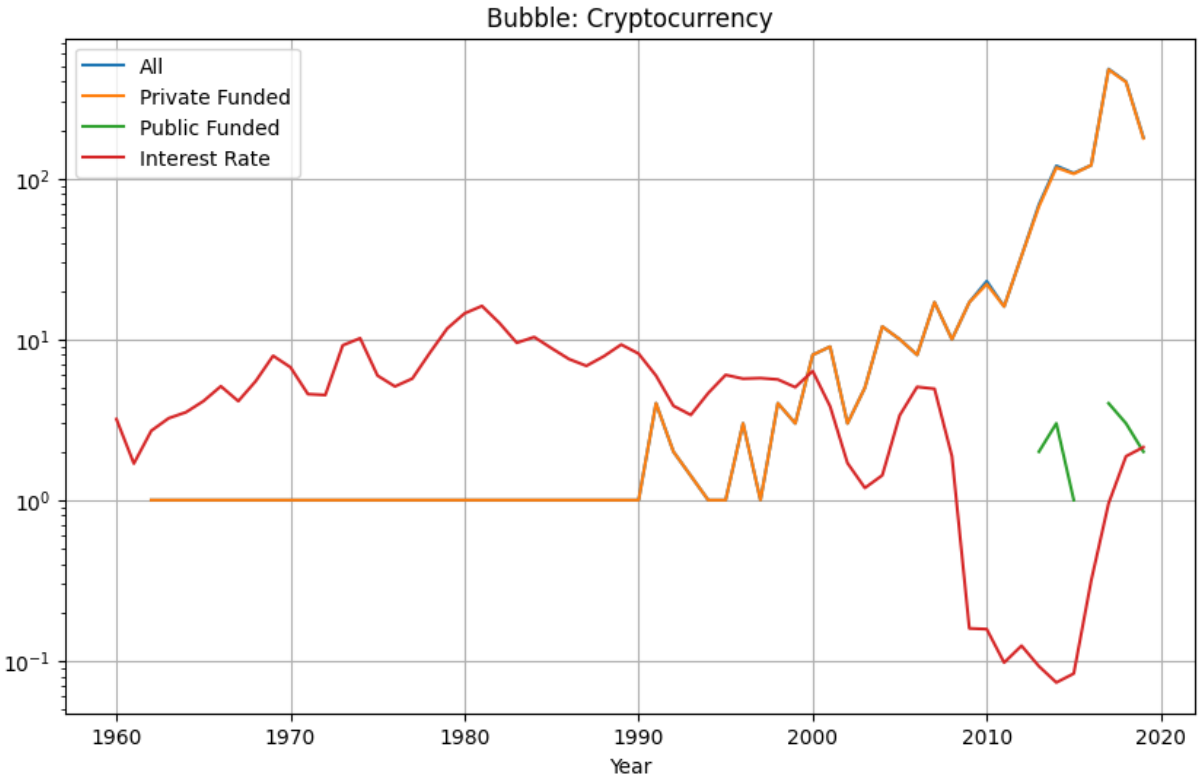

Bubble: Banking
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

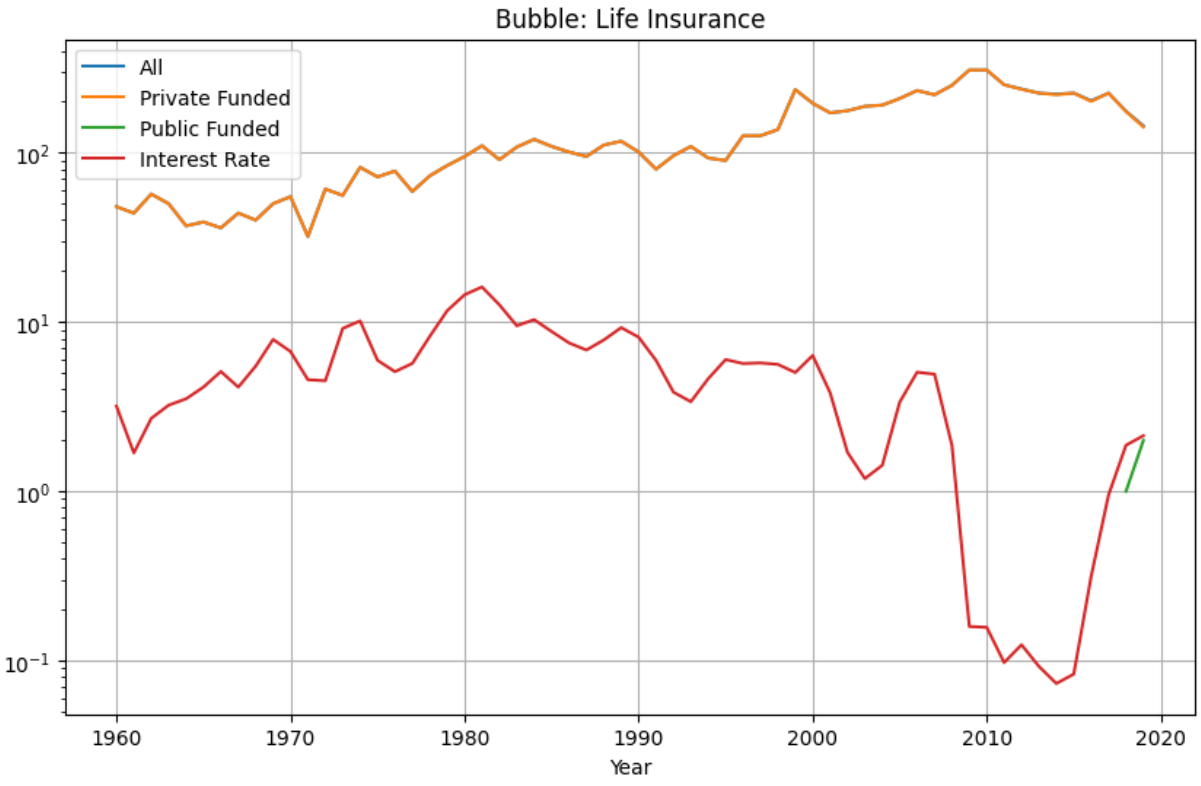

Bubble: Bitcoin
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

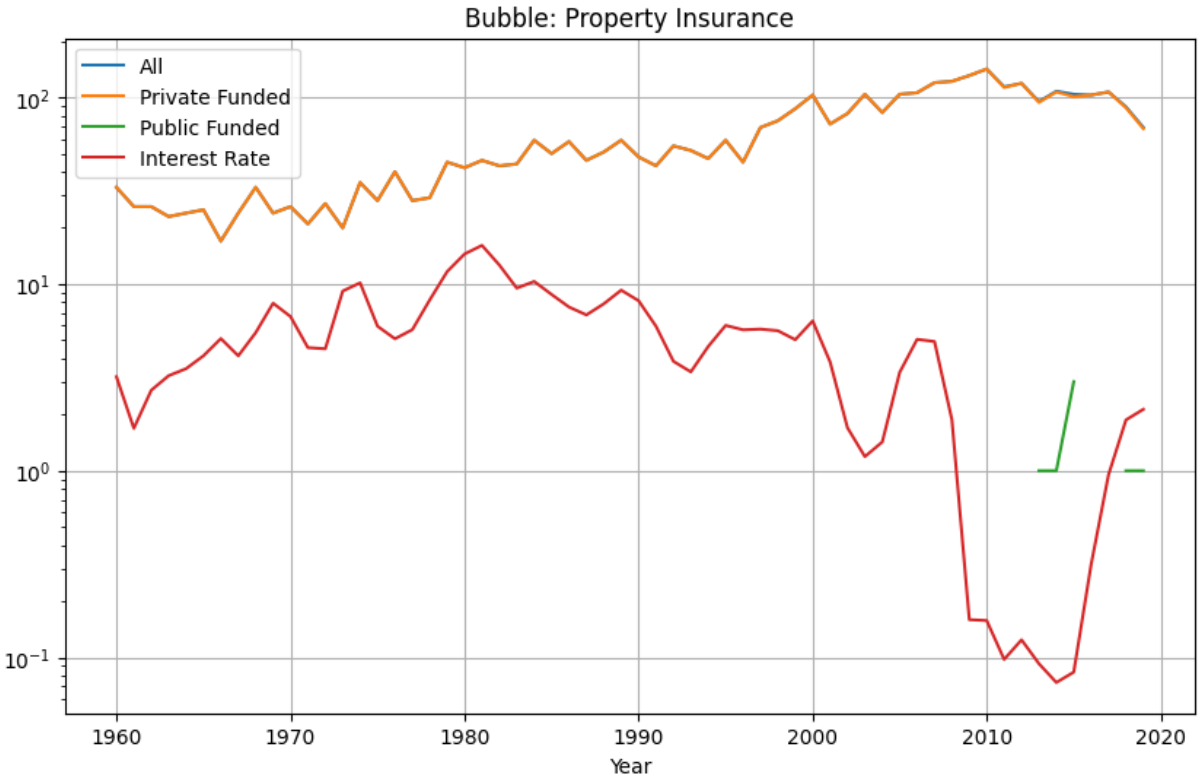

Bubble: Bookkeeping and Payroll
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year


Bubble: Commercial Insurance
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

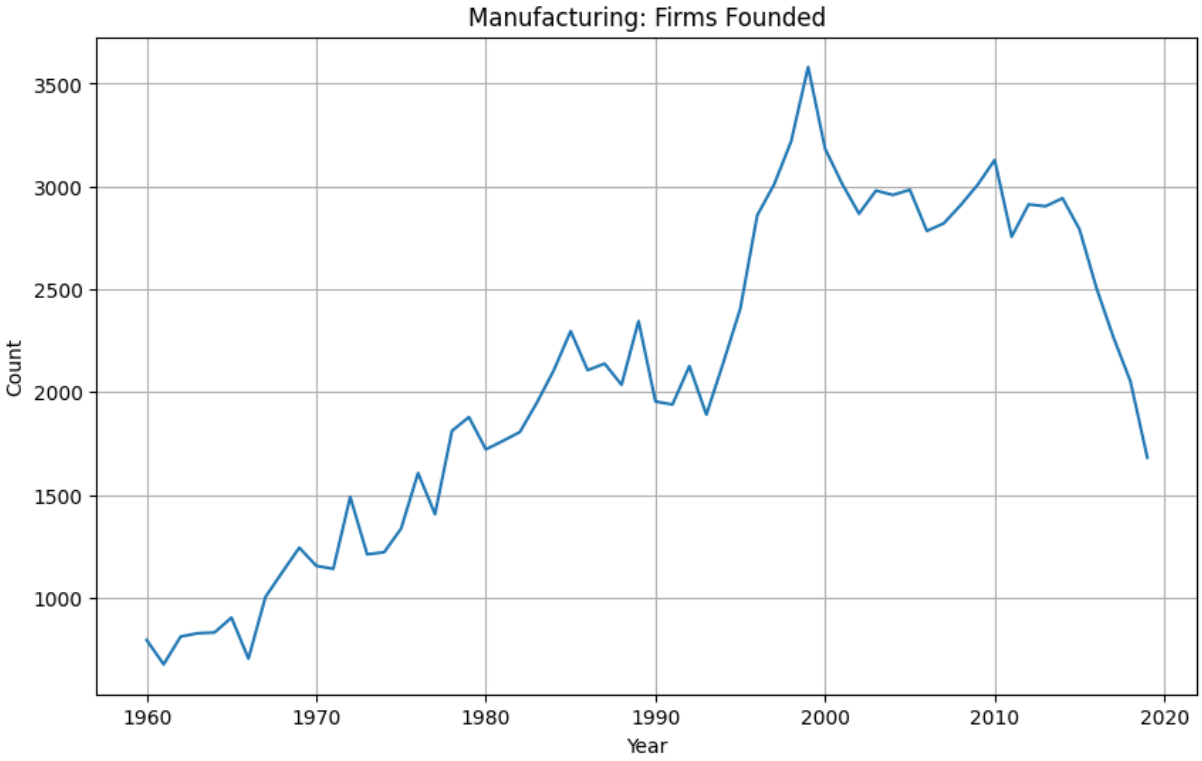

Bubble: Crowdfunding
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

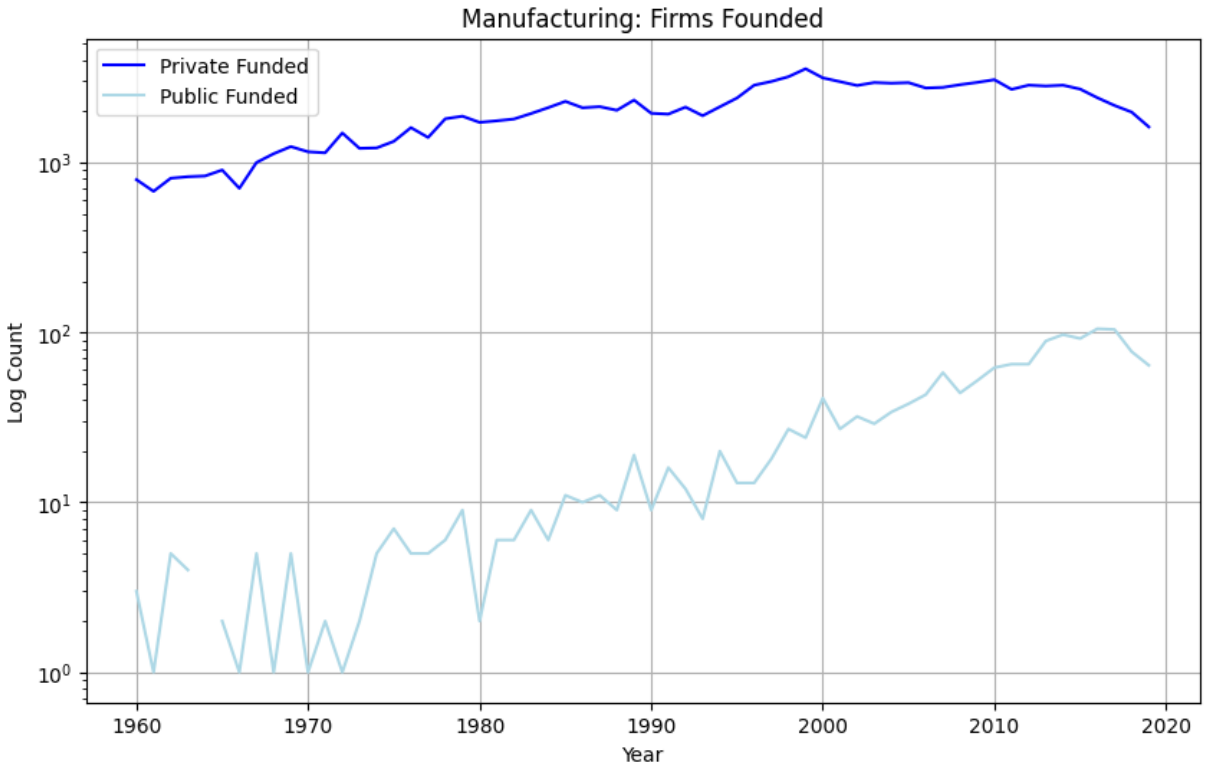

Bubble: Cryptocurrency
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

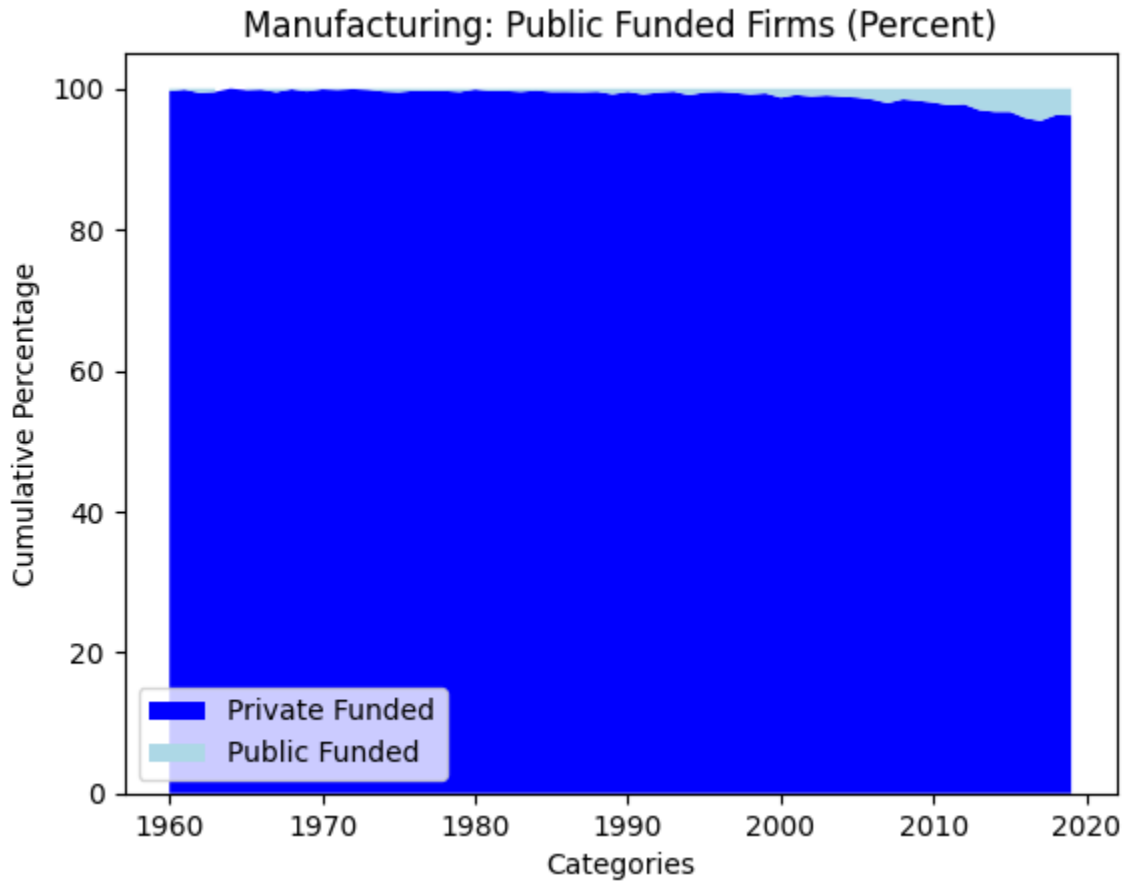

Bubble: Life Insurance
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

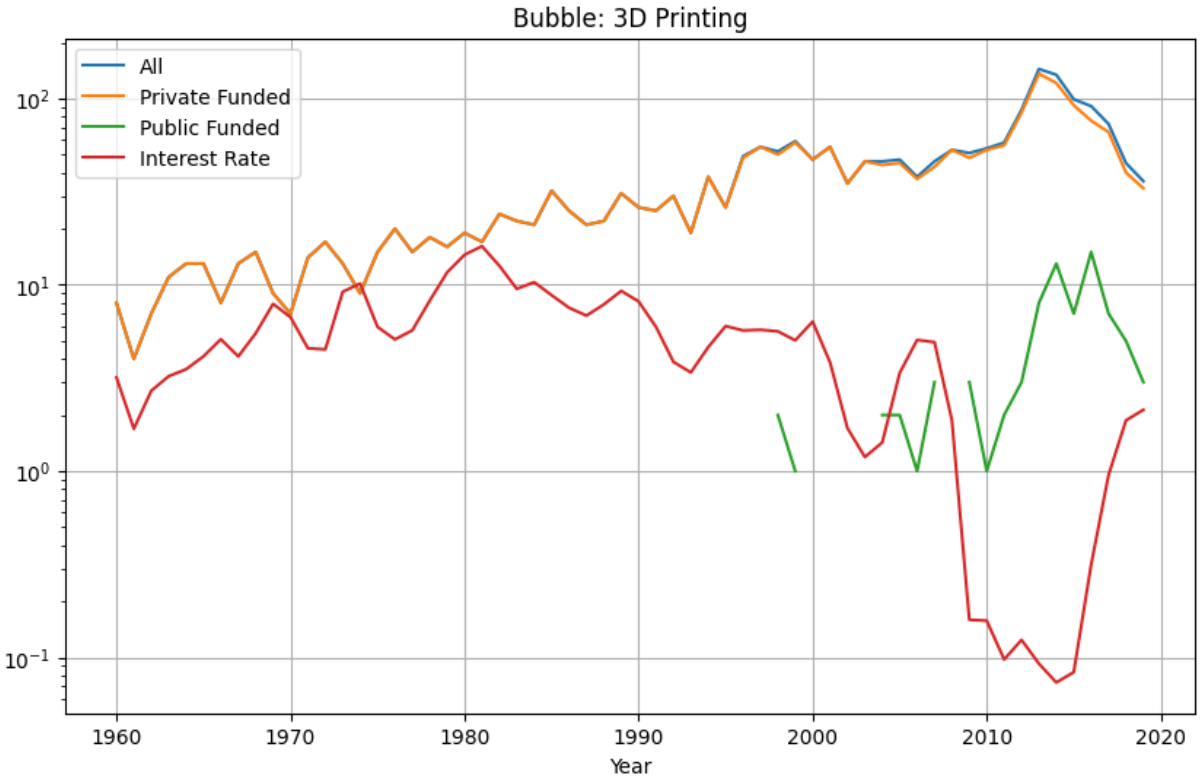


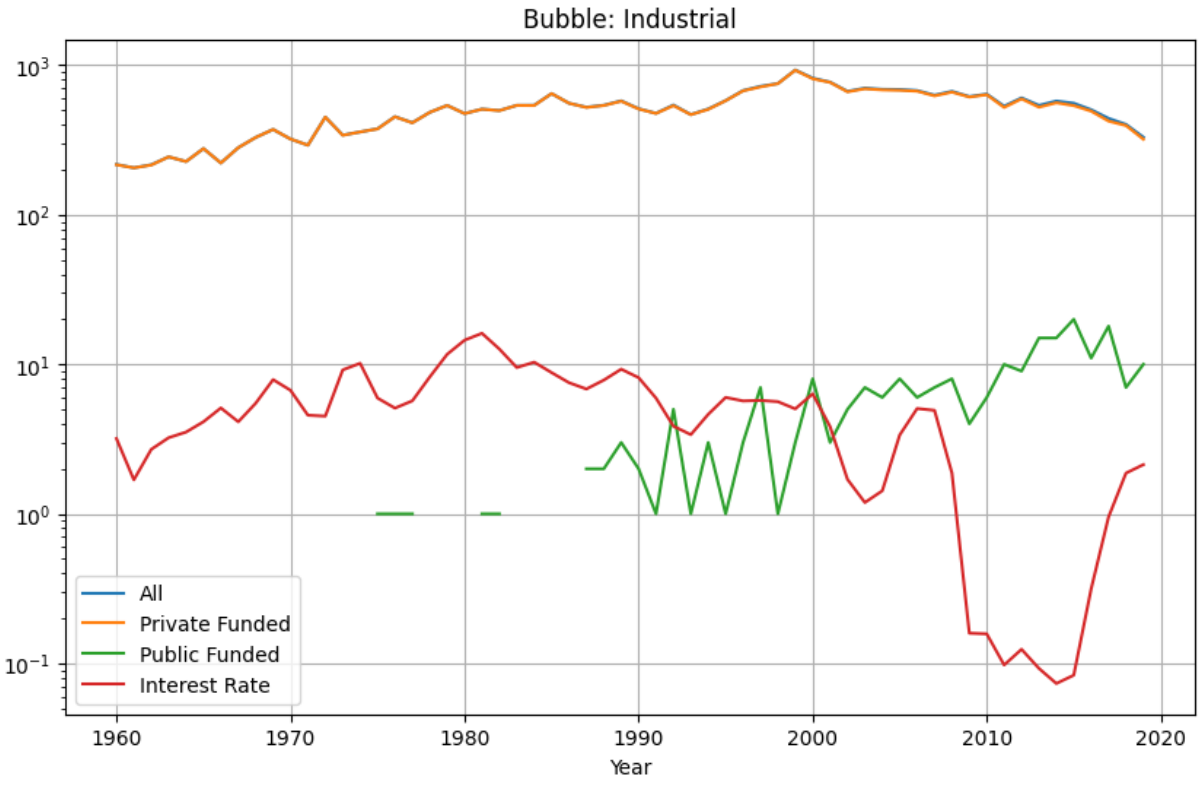
In [ ]:
```
visualise_industry_group(firms, 'Manufacturing', fed_rate)
```

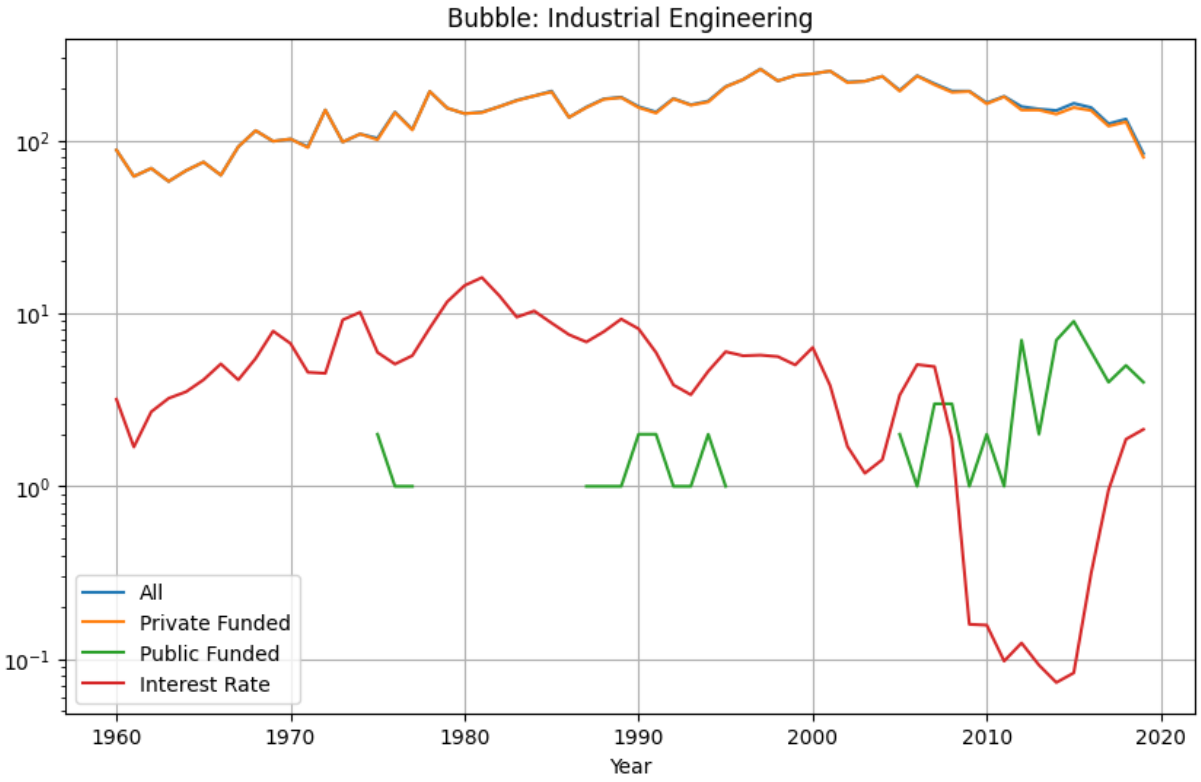


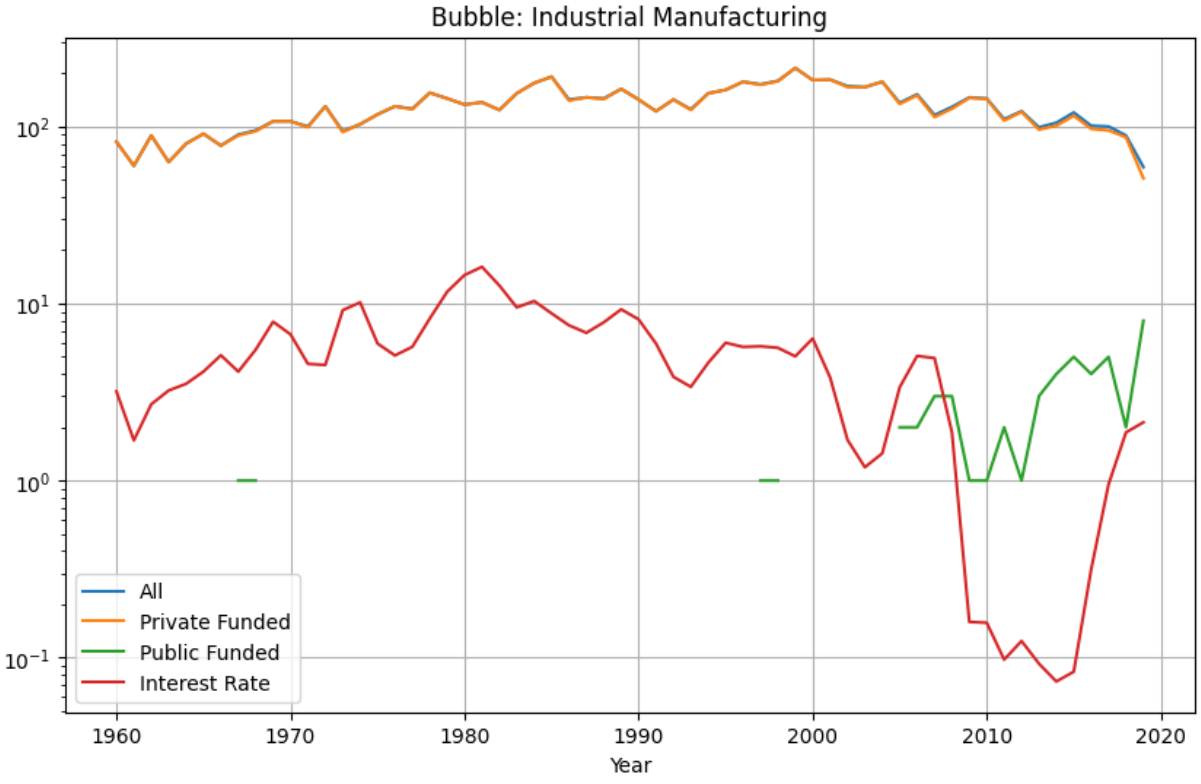

Manufacturing: Firms Founded
Private Funded
Public Funded
Log Count
10^3
10^2
10^1
10^0
1960
1970
1980
1990
2000
2010
2020
Year

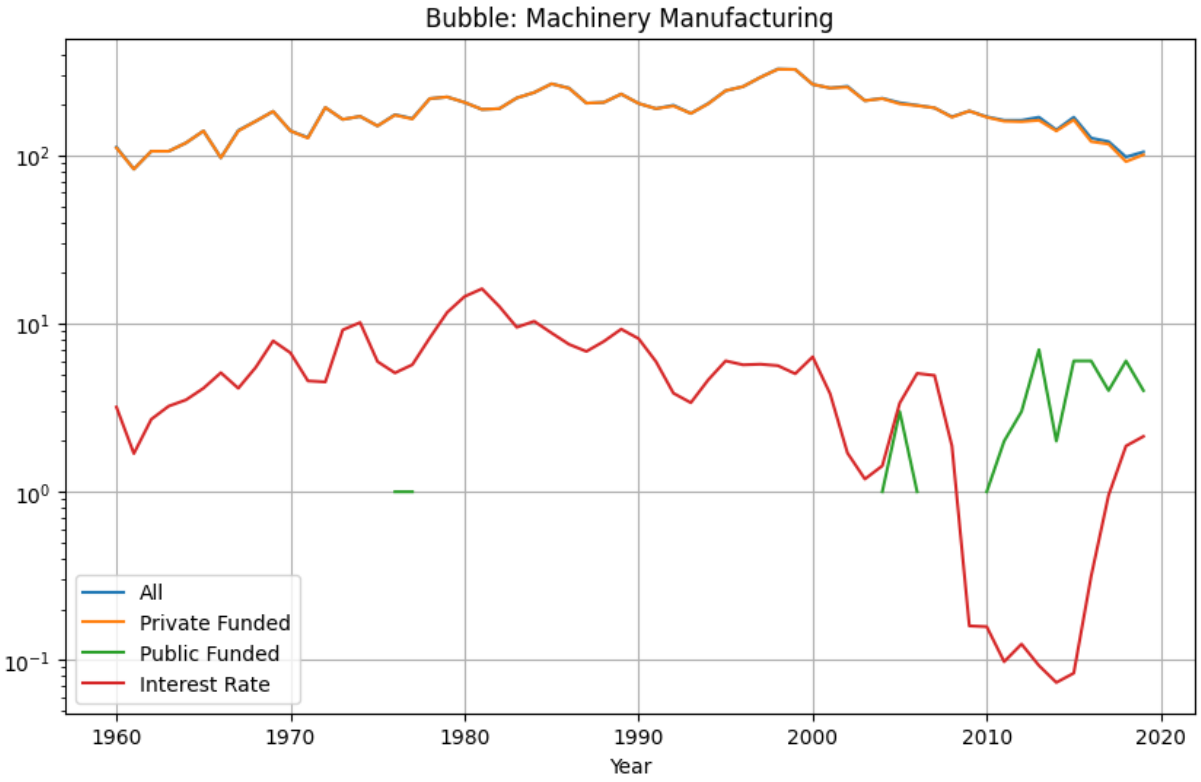

Manufacturing: Public Funded Firms (Percent)
Cumulative Percentage
100
80
60
40
20
0
Private Funded
Public Funded
1960
1970
1980
1990
2000
2010
2020
Categories

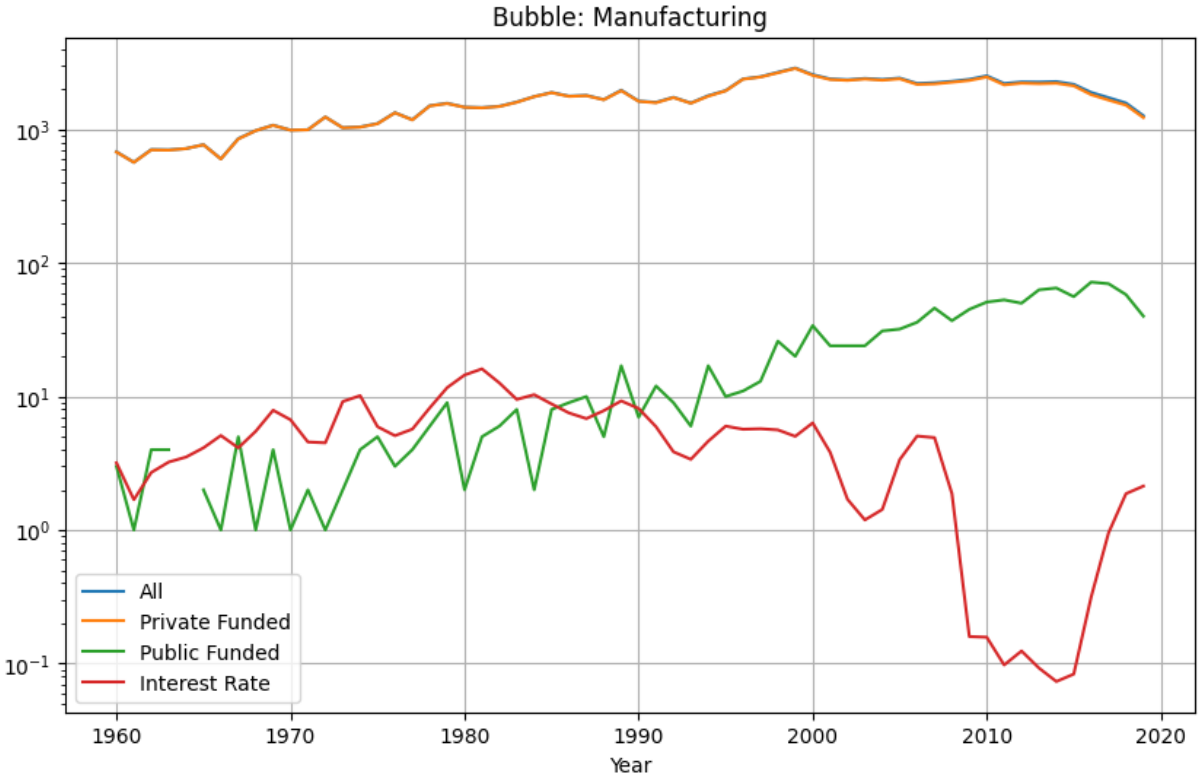

Bubble: 3D Printing
All
Private Funded
Public Funded
Interest Rate
10^2
10^1
10^0
10^{-1}
1960
1970
1980
1990
2000
2010
2020
Year

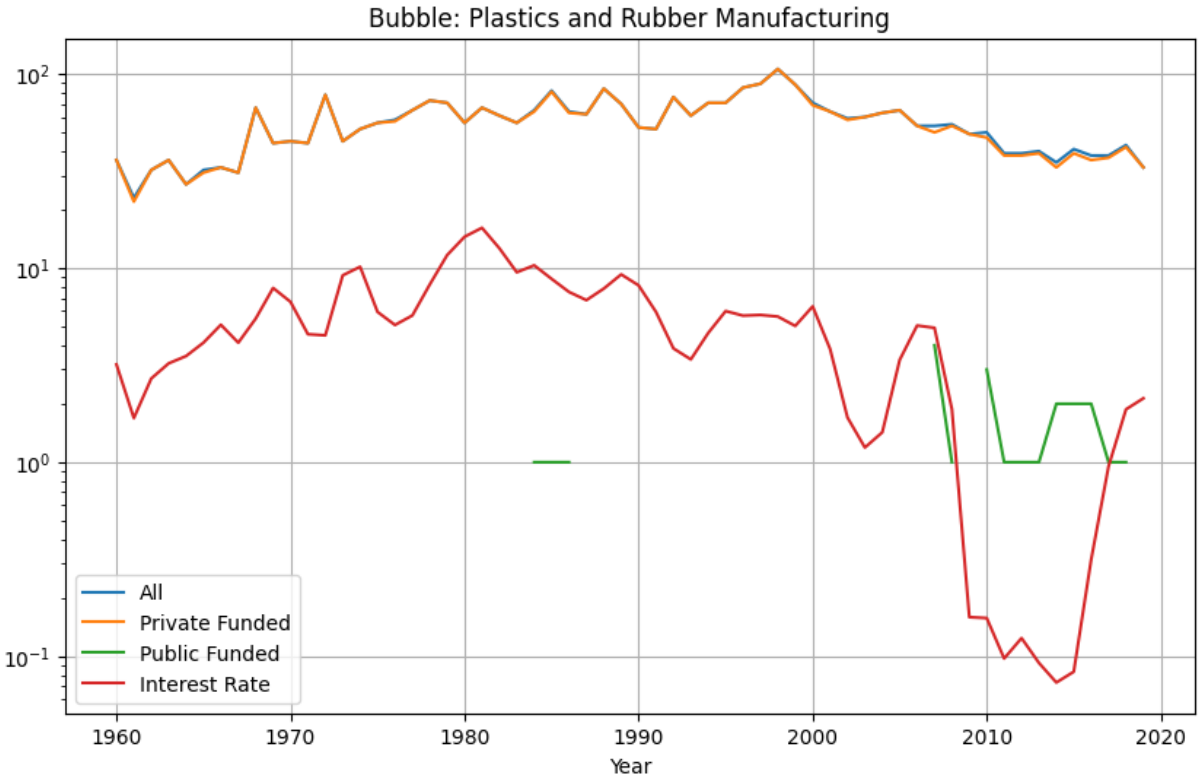

Bubble: Industrial
10^3
10^2
10^1
10^0
10^{-1}
All
Private Funded
Public Funded
Interest Rate
1960
1970
1980
1990
2000
2010
2020
Year


Bubble: Industrial Engineering
10^2
10^1
10^0
10^{-1}
All
Private Funded
Public Funded
Interest Rate
1960
1970
1980
1990
2000
2010
2020
Year


Bubble: Industrial Manufacturing
10^2
10^1
10^0
10^{-1}
All
Private Funded
Public Funded
Interest Rate
1960
1970
1980
1990
2000
2010
2020
Year


Bubble: Machinery Manufacturing
All
Private Funded
Public Funded
Interest Rate
Year


Bubble: Manufacturing
All
Private Funded
Public Funded
Interest Rate
Year


Bubble: Plastics and Rubber Manufacturing
10^2
10^1
10^0
10^{-1}
All
Private Funded
Public Funded
Interest Rate
1960
1970
1980
1990
2000
2010
2020
Year