Video Game Sales

数据展示

In [1]:

```
import pandas as pd
import numpy as np
```

```
import matplotlib.pyplot as plt
```

In [2]:

```
data = pd.read_csv('vgsales.csv')
print(data)
```

```
        Rank                                            Name Platform  \
0          1                                      Wii Sports      Wii
1          2                                Super Mario Bros.      NES
2          3                                   Mario Kart Wii      Wii
3          4                                Wii Sports Resort      Wii
4          5                          Pokemon Red/Pokemon Blue       GB
...      ...                                             ...      ...
16593  16596          Woody Woodpecker in Crazy Castle 5      GBA
16594  16597                Men in Black II: Alien Escape       GC
16595  16598  SCORE International Baja 1000: The Official Game      PS2
16596  16599                                  Know How 2       DS
16597  16600                             Spirits & Spells      GBA

         Year         Genre   Publisher  NA_Sales  EU_Sales  JP_Sales  \
0      2006.0        Sports    Nintendo     41.49     29.02      3.77
1      1985.0      Platform    Nintendo     29.08      3.58      6.81
2      2008.0        Racing    Nintendo     15.85     12.88      3.79
3      2009.0        Sports    Nintendo     15.75     11.01      3.28
4      1996.0  Role-Playing    Nintendo     11.27      8.89     10.22
...       ...           ...         ...       ...       ...       ...
16593  2002.0      Platform       Kemco      0.01      0.00      0.00
16594  2003.0       Shooter  Infogrames      0.01      0.00      0.00
16595  2008.0        Racing  Activision      0.00      0.00      0.00
16596  2010.0        Puzzle    7G//AMES      0.00      0.01      0.00
16597  2003.0      Platform     Wanadoo      0.01      0.00      0.00

       Other_Sales  Global_Sales
0             8.46         82.74
1             0.77         40.24
2             3.31         35.82
3             2.96         33.00
4             1.00         31.37
...            ...           ...
16593         0.00          0.01
16594         0.00          0.01
16595         0.00          0.01
16596         0.00          0.01
16597         0.00          0.01

[16598 rows x 11 columns]
```

受欢迎的游戏排名及前20名画图表示

In [3]:

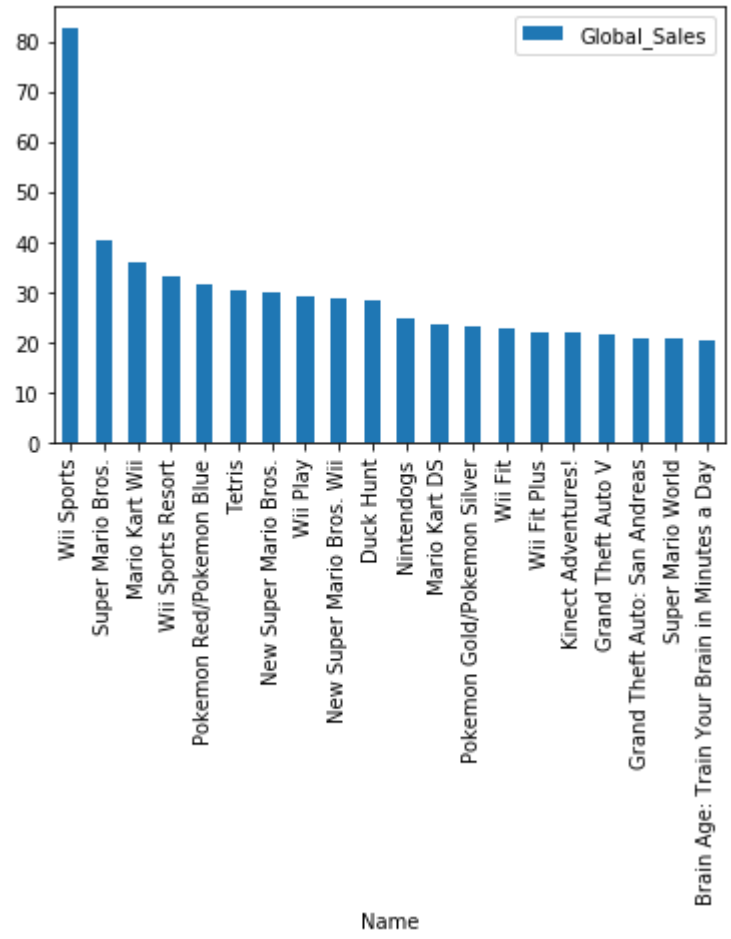
```
popular_game = data[['Name', 'Global_Sales']]
popular_game.set_index('Name', inplace=True)
print(popular_game.index)

p20_game = data[:20][['Name', 'Global_Sales']]
p20_game.set_index('Name', inplace=True)
p20_game.plot.bar()
```

```
Index(['Wii Sports', 'Super Mario Bros.', 'Mario Kart Wii',
       'Wii Sports Resort', 'Pokemon Red/Pokemon Blue', 'Tetris',
       'New Super Mario Bros.', 'Wii Play', 'New Super Mario Bros. Wii',
       'Duck Hunt',
       ...
       'Mega Brain Boost',
       'Chou Ezaru wa Akai Hana: Koi wa Tsuki ni Shirube Kareru',
       'Eiyuu Densetsu: Sora no Kiseki Material Collection Portable',
       'Myst IV: Revelation', 'Plushees', 'Woody Woodpecker in Crazy Castle 5',
       'Men in Black II: Alien Escape',
       'SCORE International Baja 1000: The Official Game', 'Know How 2',
       'Spirits & Spells'],
      dtype='object', name='Name', length=16598)
```

Out[3]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1cb2be4ab88>
```



受欢迎的游戏类型排名及前20名画图表示

In [4]:

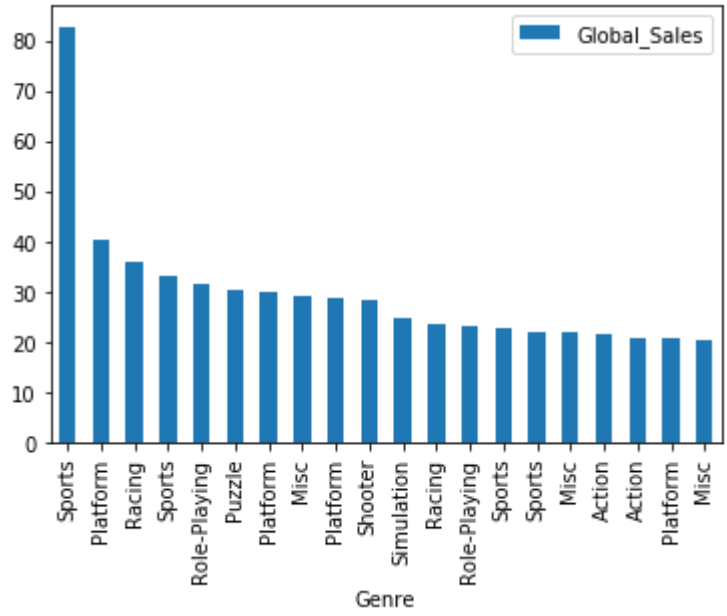
```
popular_genre = data[['Genre', 'Global_Sales']]
popular_genre.set_index('Genre', inplace=True)
print(popular_genre.index)

p20_genre = data[:20][['Genre', 'Global_Sales']]
p20_genre.set_index('Genre', inplace=True)
p20_genre.plot.bar()
```

```
Index(['Sports', 'Platform', 'Racing', 'Sports', 'Role-Playing', 'Puzzle',
       'Platform', 'Misc', 'Platform', 'Shooter',
       ...
       'Puzzle', 'Action', 'Role-Playing', 'Adventure', 'Simulation',
       'Platform', 'Shooter', 'Racing', 'Puzzle', 'Platform'],
      dtype='object', name='Genre', length=16598)
```

Out[4]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1cb2c56df08>
```



受欢迎的游戏发布平台排名及前20名画图表示

In [5]:

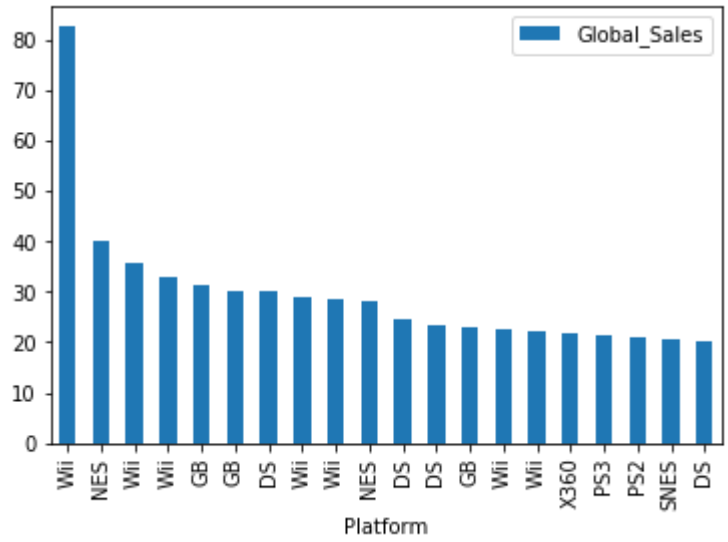
```
popular_platform = data[['Platform', 'Global_Sales']]
popular_platform.set_index('Platform', inplace=True)
print(popular_platform.index)

p20_platform = data[:20][['Platform', 'Global_Sales']]
p20_platform.set_index('Platform', inplace=True)
p20_platform.plot.bar()
```

```
Index(['Wii', 'NES', 'Wii', 'Wii', 'GB', 'GB', 'DS', 'Wii', 'Wii', 'NES',
       ...
       'DS', 'PSV', 'PSP', 'PC', 'DS', 'GBA', 'GC', 'PS2', 'DS', 'GBA'],
      dtype='object', name='Platform', length=16598)
```

Out[5]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1cb2c65ae08>
```



受欢迎的游戏发行人排名及前20名画图表示

In [6]:

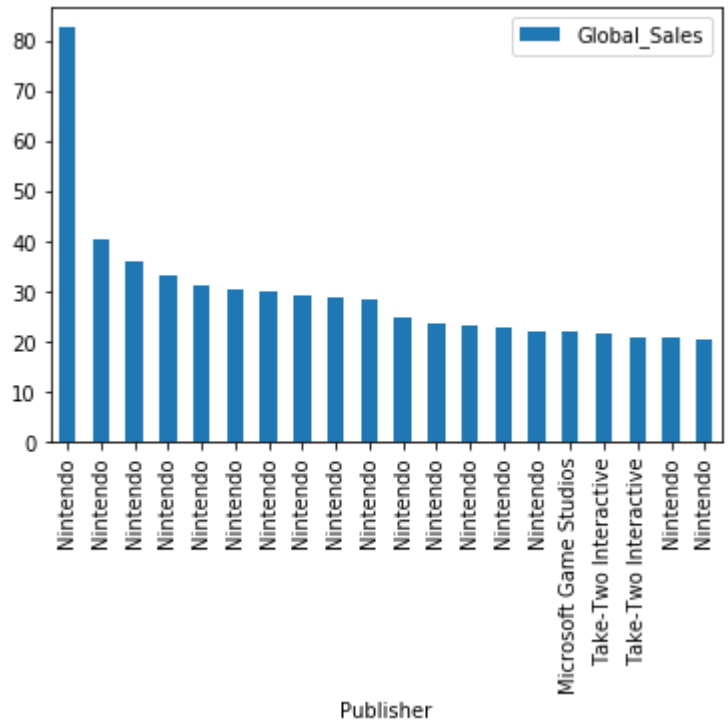
```
popular_publisher = data[['Publisher', 'Global_Sales']]
popular_publisher.set_index('Publisher', inplace=True)
print(popular_publisher.index)

p20_publisher = data[:20][['Publisher', 'Global_Sales']]
p20_publisher.set_index('Publisher', inplace=True)
p20_publisher.plot.bar()
```

```
Index([             'Nintendo',              'Nintendo',
                    'Nintendo',              'Nintendo',
                    'Nintendo',              'Nintendo',
                    'Nintendo',              'Nintendo',
                    'Nintendo',              'Nintendo',
       ...
       'Majesco Entertainment',       'dramatic create',
          'Falcom Corporation',               'Ubisoft',
                   'Destineer',                 'Kemco',
                  'Infogrames',             'Activision',
                    '7G//AMES',                'Wanadoo'],
      dtype='object', name='Publisher', length=16598)
```

Out[6]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1cb2c722d08>
```



每年的电子游戏销售额预测

In [7]:

```
#分析数据得到每年电子游戏的历史销售额数据
yearset = set()
for y in data['Year']:
    yearset.add(y)
yearlist = list(yearset)
while np.isnan(yearlist[0]):
    yearlist.remove(yearlist[0])
print(yearlist)
print(len(yearlist))

sales = []
for year in yearlist:
    year_index = data['Year'] == year
    year_salenum = data['Global_Sales'][year_index].sum()
    sales.append(year_salenum)
print(sales)
print(len(sales))
```

```
[1980.0, 1981.0, 1982.0, 1983.0, 1984.0, 1985.0, 1986.0, 1987.0, 1988.0, 1989.0, 1
990.0, 1991.0, 1992.0, 1993.0, 1994.0, 1995.0, 1996.0, 1997.0, 1998.0, 1999.0, 200
0.0, 2001.0, 2002.0, 2003.0, 2004.0, 2005.0, 2006.0, 2007.0, 2008.0, 2009.0, 2010.
0, 2011.0, 2012.0, 2013.0, 2014.0, 2015.0, 2016.0, 2017.0, 2020.0]
39
[11.38, 35.769999999999996, 28.860000000000003, 16.790000000000003, 50.36, 53.9400
00000000005, 37.06999999999999, 21.740000000000002, 47.22, 73.45000000000002, 49.3
9, 32.23, 76.16000000000003, 45.97999999999999, 79.17, 88.10999999999999, 199.1499
9999999998, 200.98, 256.47, 251.26999999999998, 201.56, 331.47, 395.5200000000000
4, 357.84999999999997, 419.30999999999995, 459.94, 521.04, 611.1300000000001, 678.
8999999999999, 667.3, 600.45, 515.99, 363.53999999999996, 368.11, 337.05, 264.44,
70.92999999999999, 0.05, 0.29]
39
```

In [8]:

```
#统计Platform、Publisher、Genre的种类
platformset = set()
for p in data['Platform']:
    platformset.add(p)
platformlist = list(platformset)
print(platformlist)
print(len(platformlist))

publisherset = set()
for pub in data['Publisher']:
    publisherset.add(pub)
publisherlist = list(publisherset)
print(publisherlist)
print(len(publisherlist))

genreset = set()
for g in data['Genre']:
    genreset.add(g)
genrelist = list(genreset)
print(genrelist)
print(len(genrelist))
```

['NG', 'N64', 'NES', 'PSV', 'PC', '3DO', 'DC', 'PSP', 'DS', 'GC', 'PS3', 'GBA', 'WS', 'XOne', 'X360', 'XB', 'PS4', 'SCD', 'PCFX', 'GB', '2600', 'Wii', 'SAT', 'Wii U', 'GEN', 'TG16', 'PS', 'PS2', '3DS', 'GG', 'SNES']
31
[nan, 'GSP', 'Epic Games', 'Agetec', 'Office Create', 'GungHo', 'Empire Interactive', 'Extreme Entertainment Group', 'Bigben Interactive', 'Piacci', 'Berkeley', 'Wargaming.net', 'Funsta', 'PM Studios', 'ESP', 'MLB.com', 'Naxat Soft', 'Laguna', 'The Learning Company', 'Vir2L Studios', 'Valve', 'Game Factory', 'Time Warner Interactive', 'Creative Core', 'BushiRoad', 'Interworks Unlimited, Inc.', 'Midway Games', 'Microprose', 'Nippon Ichi Software', 'Kando Games', 'Majesco Entertainment', 'Media Entertainment', 'Origin Systems', 'Mojang', 'Victor Interactive', 'Nexon', 'RTL', 'Level 5', 'Experience Inc.', 'Tru Blu Entertainment', '1C Company', 'Mirai Shounen', 'Jester Interactive', 'Storm City Games', 'Funcom', 'SCS Software', 'Iceberg Interactive', 'Inti Creates', 'Taito', 'Interplay', 'Merscom LLC', 'Elite', '20th Century Fox Video Games', 'Activision Value', 'Mentor Interactive', 'Brash Entertainment', 'Nippon Telenet', 'Giga', 'Moss', 'inXile Entertainment', 'Nihon Falcom Corporation', 'Tripwire Interactive', 'Mercury Games', 'Avalon Interactive', 'LEGO Media', 'Rain Games', 'Russel', 'Hip Interactive', 'Hamster Corporation', 'Marvel Entertainment', 'Irem Software Engineering', 'NCSoft', 'Tecmo Koei', 'ASCII Media Works', 'Phantom EFX', 'Natsume', 'Mumbo Jumbo', 'Magix', 'Groove Games', 'Playlogic Game Factory', 'Destination Software, Inc', 'Ackkstudios', 'Vap', 'AQ Interactive', 'Activision', 'Universal Gamex', 'Sonnet', '3DO', 'Telegames', 'Glams', 'Destineer', 'KSS', 'Kool Kizz', 'BMG Interactive Entertainment', 'Licensed 4U', 'Imax', 'Kalypso Media', 'Lexicon Entertainment', 'Altron', 'DreamWorks Interactive', 'FuRyu Corporation', 'Kadokawa Shoten', 'Seta Corporation', 'Sony Music Entertainment', 'Lighthouse Interactive', 'Accolade', 'Jaleco', 'DHM Interactive', 'Sony Online Entertainment', 'Atari', 'Kamui', 'Nordcurrent', 'Xing Entertainment', 'GOA', 'Zoo Games', 'Ghostlight', 'Illusion Softworks', 'NetRevo', 'Asmik Corp', 'Elf', 'Data East', 'Xseed Games', 'Societa', 'HMH Interactive', 'Intergrow', 'Rebellion Developments', 'Gameloft', 'ASK', 'TechnoSoft', 'Hasbro Interactive', 'Colecо', '505 Games', 'Compile', 'Slightly Mad Studios', 'dramatic create', 'U.S. Gold', 'Men-A-Vision', 'Phenomedia', 'NovaLogic', 'Sears', 'Revolution (Japan)', 'Imagineer', 'Foreign Media Games', 'Avanquest', 'Simon & Schuster Interactive', 'iWin', 'Sammy Corporation', 'Arena Entertainment', 'Core Design Ltd.', 'FunSoft', 'NCS', 'Stainless Games', 'Essential Games', 'Harmonix Music Systems', 'fonfun', 'Konami Digital Entertainment', 'Number None', 'Agatsuma Entertainment', 'Culture Publishers', 'Westwood Studios', 'Rondomedia', 'O-Games', 'Spike', 'Arc System Works', 'Imadio', 'Nippon Amuse', 'Codemasters', 'Midas Interactive Entertainment', 'id Software', 'Black Bean Games', 'Mastiff', 'Quest', 'Giza10', 'Hudson Entertainment', 'Performance Designed Products', 'Avanquest Software', 'Big Fish Games', '989 Studios', 'Deep Silver', 'Parker Bros.', 'Pow', 'Hudson Soft', 'Sony Computer Entertainment America', 'Michaelsoft', 'NDA Productions', 'Rising Star Games', 'Mud Duck Productions', 'Touchstone', 'General Entertainment', 'Gainax Network Systems', 'TopWare Interactive', 'Red Orb', 'On Demand', 'Paon', 'Myelin Media', 'Reef Entertainment', 'Otomate', 'Kadokawa Games', 'Broccoli', 'Hearty Robin', 'Genki', 'LSP Games', 'SquareSoft', 'MC2 Entertainment', 'JoWood Productions', 'Blast! Entertainment Ltd', 'Marvelous Games', 'ASC Games', 'Sold Out', 'Comfort', 'Vivendi Games', 'TOHO', 'Liquid Games', 'Graffiti', "Yuke's", 'Codemasters Online', '49Games', 'Epoch', 'Fields', 'Angel Studios', 'Titus', 'Focus Multimedia', '989 Sports', 'Aqua Plus', 'Asgard', 'Black Label Games', 'Sunflowers', 'PopTop Software', 'Nobilis', 'Interchannel-Holon', 'Riverhillsoft', 'Plenty', 'Electronic Arts', 'Magical Company', 'Home Entertainment Suppliers', 'RedOctane', 'Havas Interactive', 'Zoo Digital Publishing', 'Universal Interactive', 'Nippon Columbia', 'Vic Tokai', 'TGL', 'Saurus', 'UIG Entertainment', 'Enjoy Gaming ltd.', 'Mycom', 'Media Works', 'HuneX', 'Mitsui', '2D Boy', 'Milestone', 'Enix Corporation', 'Jack of All Games', 'System 3 Arcade Software', 'Valve Software', 'Funbox Media', 'Encore', 'CokeM Interactive', 'Axela', 'Cloud Imperium Games Corporation', 'Sunrise Interactive', '7G//AMES', 'Coconuts Japan', 'Minato Station', 'Panther Software', 'SNK', 'Her Interactive', 'Bomb', 'Crystal Dynamics', 'Locus', 'Video System', 'Telstar', 'Views', 'Kokopeli Digital Studios', 'GN Software', 'Gakken', 'Fuji', 'JVC', 'Psygnosis', 'Aria', 'Alternative Software', 'Rocket Company', 'Enterbrain', 'Hello Games', 'SPS', 'Ocean',

'Pioneer LDC', 'Capcom', 'Interchannel', 'Boost On', 'Knowledge Adventure', 'Adeli
ne Software', 'Quinrose', 'Paon Corporation', 'Daedalic Entertainment', 'Playmate
s', 'Telltale Games', 'Devolver Digital', 'Data Design Interactive', 'T&E Soft',
'Takuyo', 'Rebellion', 'TDK Mediactive', 'Image Epoch', 'Cave', 'Quintet', 'SCi',
'Answer Software', 'Idea Factory', 'Square EA', 'CTO SpA', 'Fortyfive', 'ChunSof
t', 'Sega', 'Infogrames', 'Virgin Interactive', 'DSI Games', 'Warp', 'Vatical Ente
rtainment', 'Dorart', 'Zenrin', 'Summitsoft', 'Electronic Arts Victor', 'Abyligh
t', 'PQube', 'Yumedia', 'Mamba Games', 'Gamecock', 'Tryfirst', 'New World Computin
g', 'Mastertronic', 'Compile Heart', 'Paradox Interactive', '10TACLE Studios', 'Fl
ashpoint Games', 'New', 'Ertain', 'Ultravision', 'Aques', 'Ivolgamus', 'Acclaim En
tertainment', 'Monte Christo Multimedia', 'Alvion', 'Arika', 'Navarre Corp', 'Maxi
mum Family Games', 'Pacific Century Cyber Works', 'Aksys Games', 'Ongakukan', 'New
KidCo', 'Excalibur Publishing', 'Asmik Ace Entertainment', 'Gust', 'Play It', 'Dae
dalic', 'King Records', 'Syscom', 'Athena', 'Conspiracy Entertainment', 'Trion Wor
lds', 'G.Rev', 'Kemco', 'Oxygen Interactive', 'Seventh Chord', 'City Interactive',
'Yacht Club Games', 'Jorudan', 'Culture Brain', 'Disney Interactive Studios', 'Scr
eenlife', 'Headup Games', 'Global Star', 'BPS', 'bitComposer Games', 'Tradewest',
'Acquire', 'Technos Japan Corporation', 'Nichibutsu', 'Micro Cabin', 'Mad Catz',
'Revolution Software', 'Introversion Software', 'Warner Bros. Interactive Entertai
nment', 'Red Storm Entertainment', 'Gaga', 'Wanadoo', 'Edia', 'Crimson Cow', 'Cybe
rFront', 'THQ', 'Namco Bandai Games', 'Mattel Interactive', 'Princess Soft', 'Luca
sArts', 'Human Entertainment', 'Astragon', 'Yeti', 'Square', 'Game Life', 'Banpres
to', 'Little Orbit', 'HAL Laboratory', 'Sunsoft', 'Indie Games', 'ITT Family Game
s', 'imageepoch Inc.', 'MTV Games', 'Dusenberry Martin Racing', 'Detn8 Games', 'Mi
lestone S.r.l.', 'Data Age', 'BAM! Entertainment', 'Starfish', 'CDV Software Enter
tainment', 'D3Publisher', 'Xicat Interactive', 'TDK Core', 'Popcorn Arcade', 'Happ
inet', 'Alchemist', 'Yamasa Entertainment', 'NEC Interchannel', 'Maxis', 'Zushi Ga
mes', 'FuRyu', 'Atlus', 'WayForward Technologies', 'Interplay Productions', 'From
Software', 'Global A Entertainment', 'Playmore', 'Xplosiv', 'Phantagram', 'UFO Int
eractive', 'Shogakukan', 'Asylum Entertainment', 'Grand Prix Games', 'Evolved Game
s', 'Milestone S.r.l', 'P2 Games', 'Nordic Games', '5pb', 'mixi, Inc', 'Graphsim E
ntertainment', 'Ubisoft Annecy', 'Daito', 'Kids Station', 'American Softworks', 'A
lawar Entertainment', 'Gamebridge', 'Aspyr', 'Scholastic Inc.', 'MediaQuest', 'Hac
kberry', 'Takara Tomy', 'Wizard Video Games', 'Palcom', 'Aruze Corp', 'Imagework
s', 'Mystique', 'DigiCube', 'Metro 3D', 'Evolution Games', 'Ascaron Entertainmen
t', 'Tetris Online', 'Phoenix Games', 'GT Interactive', 'MTO', 'KID', 'responDESIG
N', 'Mindscape', 'Datam Polystar', 'Karin Entertainment', 'Valcon Games', 'Pack-In
-Video', '03 Entertainment', 'Microids', 'ASCII Entertainment', 'Bohemia Interacti
ve', 'Starpath Corp.', 'ArtDink', 'PlayV', 'EON Digital Entertainment', 'Microsoft
Game Studios', 'Takara', 'System Soft', 'Take-Two Interactive', 'System 3', 'Bethe
sda Softworks', 'Ubisoft', 'SSI', 'Kaga Create', 'CCP', 'Gathering of Developers',
'SNK Playmore', 'PopCap Games', 'Crave Entertainment', 'DTP Entertainment', 'Virtu
al Play Games', 'Sweets', 'Ignition Entertainment', 'Focus Home Interactive', 'Typ
e-Moon', 'Cygames', 'Blue Byte', 'Gremlin Interactive Ltd', 'Warashi', 'NEC', 'Jus
t Flight', 'EA Games', 'Easy Interactive', 'Nitroplus', 'Team17 Software', 'Unknow
n', 'White Park Bay Software', 'Max Five', 'Paradox Development', 'CBS Electronic
s', 'GameTek', 'Quelle', 'Genterprise', 'Sony Computer Entertainment', 'Tigervisio
n', 'Visco', 'Game Arts', 'Aerosoft', 'Activision Blizzard', 'TYO', 'Legacy Intera
ctive', 'Misawa', 'GameMill Entertainment', 'Imagic', 'Strategy First', 'Tommo',
'Eidos Interactive', 'Idea Factory International', 'IE Institute', 'Sting', 'Benes
se', 'Pony Canyon', 'Sony Computer Entertainment Europe', 'Gotham Games', 'Fox Int
eractive', 'Prototype', 'SouthPeak Games', 'Swing! Entertainment', 'Marvelous Ente
rtainment', 'Commseed', 'Square Enix', 'UEP Systems', 'Insomniac Games', 'Hect',
'Flight-Plan', 'Koch Media', 'Big Ben Interactive', 'ValuSoft', 'Griffin Internati
onal', 'Marvelous Interactive', 'Tivola', 'DreamCatcher Interactive', 'Success',
'RED Entertainment', 'Rage Software', 'Nintendo', 'Masque Publishing', 'Pinnacle',
'CPG Products', 'Tomy Corporation', 'XS Games', 'Slitherine Software', 'Media Ring
s', 'Ecole', 'TalonSoft', 'Falcom Corporation', 'Media Factory', 'Ascaron Entertai
nment GmbH', 'Pack In Soft', 'Neko Entertainment', 'The Adventure Company']
579
['Puzzle', 'Action', 'Strategy', 'Fighting', 'Simulation', 'Sports', 'Role-Playin

g', 'Racing', 'Misc', 'Adventure', 'Shooter', 'Platform']
12

In [9]:

```
#选择排名前20的platform、publisher和12种Genre作为自变量

genre_data = data[['Genre','Global_Sales']]
genre_data.set_index('Genre',inplace=True)
genre_data_20 = genre_data.groupby('Genre').sum()
genre_data_20.sort_values(by='Global_Sales', ascending=False, inplace=True)
genre_data_20 = genre_data_20[:20]
#print(genre_data_20)

platform_data = data[['Platform','Global_Sales']]
platform_data.set_index('Platform',inplace=True)
platform_data_20 = platform_data.groupby('Platform').sum()
platform_data_20.sort_values(by='Global_Sales', ascending=False, inplace=True)
platform_data_20 = platform_data_20[:20]
#print(platform_data_20)

publisher_data = data[['Publisher','Global_Sales']]
publisher_data.set_index('Publisher',inplace=True)
publisher_data_20 = publisher_data.groupby('Publisher').sum()
publisher_data_20.sort_values(by='Global_Sales', ascending=False, inplace=True)
publisher_data_20 = publisher_data_20[:20]
#print(publisher_data_20)

score = 0
Inputvec = {}
for item in genre_data_20.index:
        Inputvec[item] = score
        score += 1
for item in platform_data_20.index:
        Inputvec[item] = score
        score += 1
for item in publisher_data_20.index:
        Inputvec[item] = score
        score += 1
print(Inputvec)

ys = 0
yearvec = {}
for y in yearlist:
    yearvec[y] = ys
    ys +=1
print(yearvec)
```

{'Action': 0, 'Sports': 1, 'Shooter': 2, 'Role-Playing': 3, 'Platform': 4, 'Misc': 5, 'Racing': 6, 'Fighting': 7, 'Simulation': 8, 'Puzzle': 9, 'Adventure': 10, 'Strategy': 11, 'PS2': 12, 'X360': 13, 'PS3': 14, 'Wii': 15, 'DS': 16, 'PS': 17, 'GBA': 18, 'PSP': 19, 'PS4': 20, 'PC': 21, 'XB': 22, 'GB': 23, 'NES': 24, '3DS': 25, 'N64': 26, 'SNES': 27, 'GC': 28, 'XOne': 29, '2600': 30, 'WiiU': 31, 'Nintendo': 32, 'Electronic Arts': 33, 'Activision': 34, 'Sony Computer Entertainment': 35, 'Ubisoft': 36, 'Take-Two Interactive': 37, 'THQ': 38, 'Konami Digital Entertainment': 39, 'Sega': 40, 'Namco Bandai Games': 41, 'Microsoft Game Studios': 42, 'Capcom': 43, 'Atari': 44, 'Warner Bros. Interactive Entertainment': 45, 'Square Enix': 46, 'Disney Interactive Studios': 47, 'Eidos Interactive': 48, 'LucasArts': 49, 'Bethesda Softworks': 50, 'Midway Games': 51}
{1980.0: 0, 1981.0: 1, 1982.0: 2, 1983.0: 3, 1984.0: 4, 1985.0: 5, 1986.0: 6, 1987.0: 7, 1988.0: 8, 1989.0: 9, 1990.0: 10, 1991.0: 11, 1992.0: 12, 1993.0: 13, 1994.0: 14, 1995.0: 15, 1996.0: 16, 1997.0: 17, 1998.0: 18, 1999.0: 19, 2000.0: 20, 2001.0: 21, 2002.0: 22, 2003.0: 23, 2004.0: 24, 2005.0: 25, 2006.0: 26, 2007.0: 27, 2008.0: 28, 2009.0: 29, 2010.0: 30, 2011.0: 31, 2012.0: 32, 2013.0: 33, 2014.0: 34, 2015.0: 35, 2016.0: 36, 2017.0: 37, 2020.0: 38}

In [10]:

```
#生成输入向量
index = ['Genre', 'Platform', 'Publisher']
Input = np.zeros((len(yearlist), 52))

#genre
data1 = data[['Year','Genre','Global_Sales']]
#data1.set_index(['Year','Genre'],inplace=True)
for ye in yearlist:
    for gen in genre_data_20.index:
        df = data1[(data1['Year']==ye) & (data1['Genre']==gen)]
        #print(df)
        #print(df['Global_Sales'].sum())
        numsum = df['Global_Sales'].sum()
        a = yearvec[ye]
        b = Inputvec[gen]
        Input[a][b] = numsum

#platform
data2 = data[['Year','Platform','Global_Sales']]
for ye in yearlist:
    for pla in platform_data_20.index:
        df2 = data2[(data2['Year']==ye) & (data2['Platform']==pla)]
        #print(df)
        #print(df['Global_Sales'].sum())
        numsum = df2['Global_Sales'].sum()
        a = yearvec[ye]
        b = Inputvec[pla]
        Input[a][b] = numsum

#publisher
data3 = data[['Year','Publisher','Global_Sales']]
for ye in yearlist:
    for pub in publisher_data_20.index:
        df3 = data3[(data3['Year']==ye) & (data3['Publisher']==pub)]
        #print(df)
        #print(df['Global_Sales'].sum())
        numsum = df3['Global_Sales'].sum()
        a = yearvec[ye]
        b = Inputvec[pub]
        Input[a][b] = numsum

print(Input)
```

```
[[3.400e-01 4.900e-01 7.070e+00 ... 0.000e+00 0.000e+00 0.000e+00]
 [1.484e+01 7.900e-01 1.004e+01 ... 0.000e+00 0.000e+00 0.000e+00]
 [6.520e+00 1.050e+00 3.790e+00 ... 0.000e+00 0.000e+00 0.000e+00]
 ...
 [1.991e+01 1.460e+01 1.822e+01 ... 0.000e+00 2.250e+00 0.000e+00]
 [1.000e-02 0.000e+00 0.000e+00 ... 0.000e+00 0.000e+00 0.000e+00]
 [0.000e+00 0.000e+00 0.000e+00 ... 0.000e+00 0.000e+00 0.000e+00]]
```

In [11]:

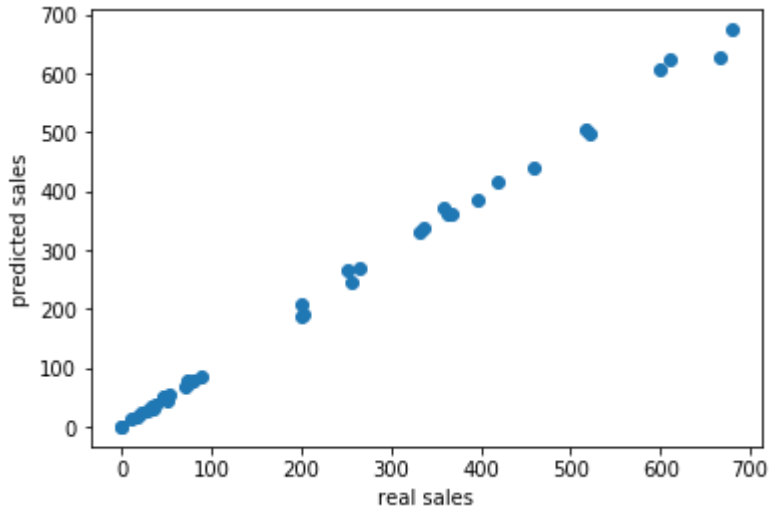
```
#线性回归预测
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import KFold
x = Input
y = np.array(sales)

from scipy.stats import pearsonr
x = x  / np.sum(x, 0)
print(x)
print(x.shape)
print(y.shape)
kf = KFold(n_splits=5, shuffle=True, random_state=123)
model = LinearRegression()

y_pred = np.zeros(y.shape)
for train, test in kf.split(x, y):
    model.fit(x[train], y[train])
    y_pred[test] = model.predict(x[test])

rmse = np.sqrt(np.sum((y_pred - y) ** 2) / y.shape[0])
print(rmse)
r = pearsonr(y, y_pred)[0]
print(r)
plt.scatter(y, y_pred)
plt.xlabel('real sales')
plt.ylabel('predicted sales')
plt.show()
```

```
[[1.97343982e-04 3.74262931e-04 6.88949523e-03 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 [8.61348440e-03 6.03403501e-04 9.78366790e-03 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 [3.78436107e-03 8.01991995e-04 3.69323719e-03 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 ...
 [1.15562314e-02 1.11515077e-02 1.77548236e-02 ... 0.00000000e+00
  2.75972035e-02 0.00000000e+00]
 [5.80423477e-06 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 ... 0.00000000e+00
  0.00000000e+00 0.00000000e+00]]
(39, 52)
(39,)
10.329579591583872
0.998965718955179
```

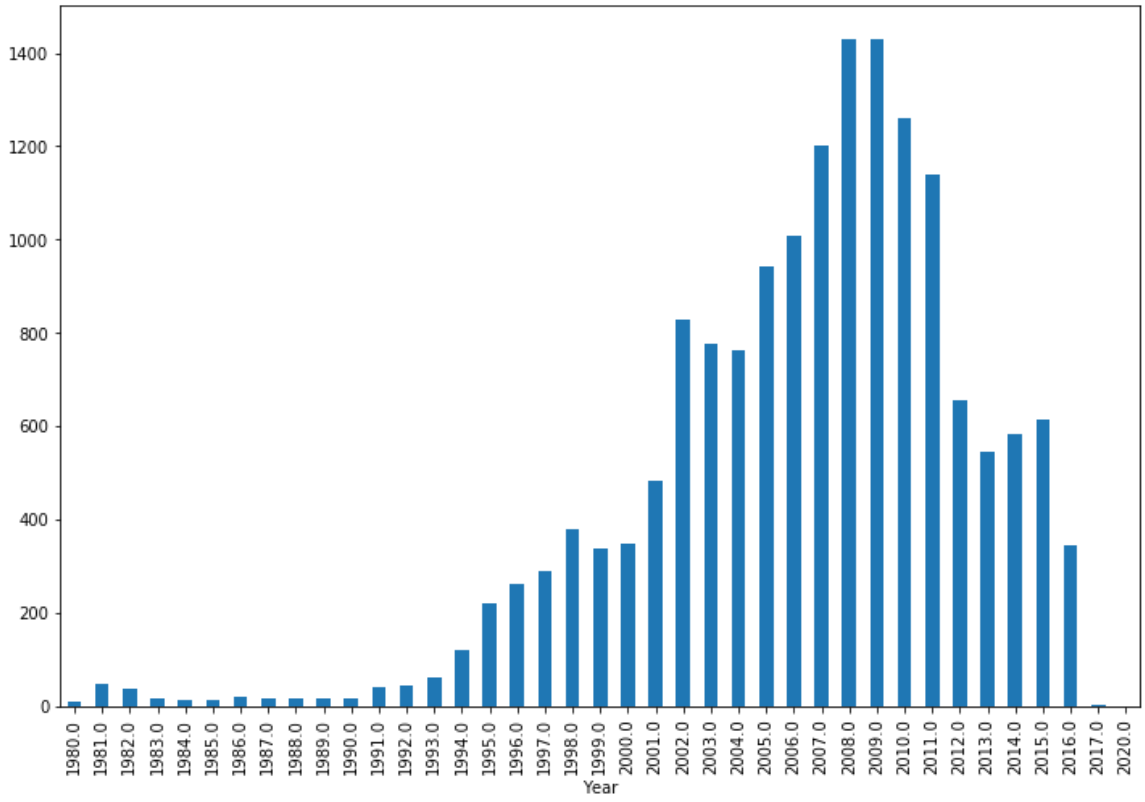


可视化应用，描述销售概况

每年上线的游戏数量

In [12]:

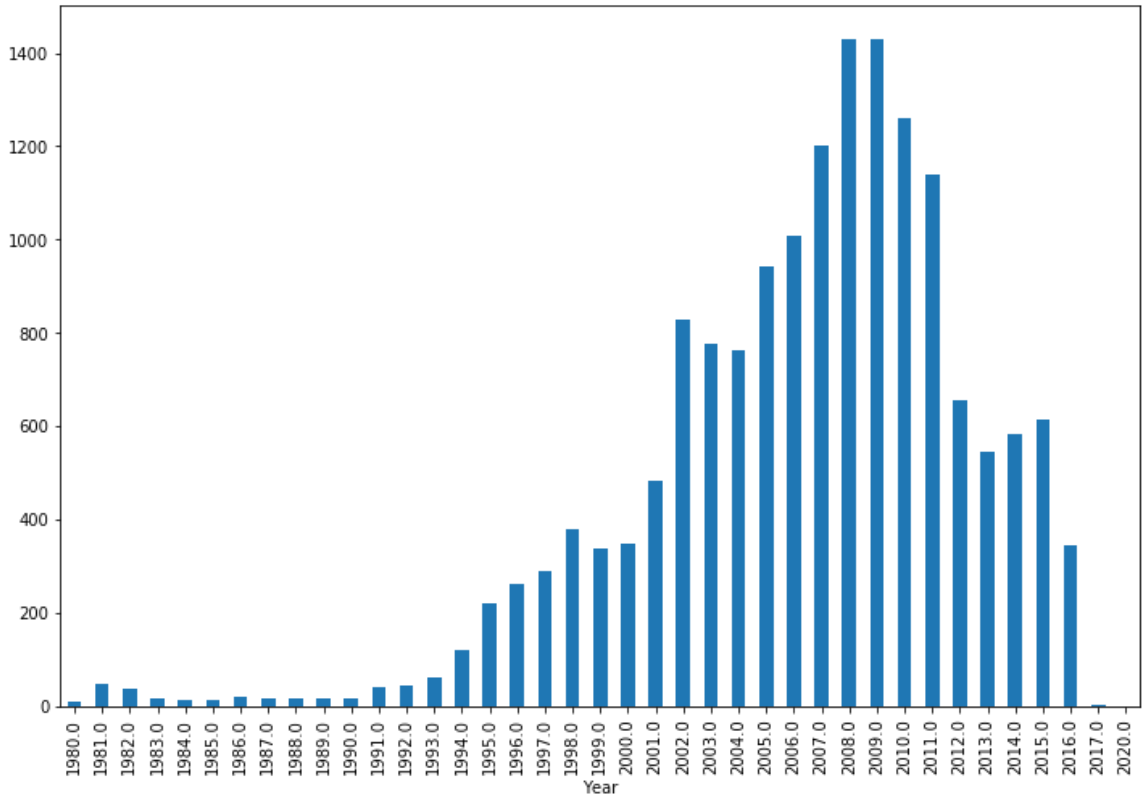
```
x = data.groupby(['Year']).count()
gamenum = x['Name']
plt.figure(figsize=(12,8))
gamenum.plot.bar()
plt.show()
#由图中可见，游戏的发行量呈现先增后减的状态，在2008、2009年达到峰值
```



每年的总销售量

In [13]:

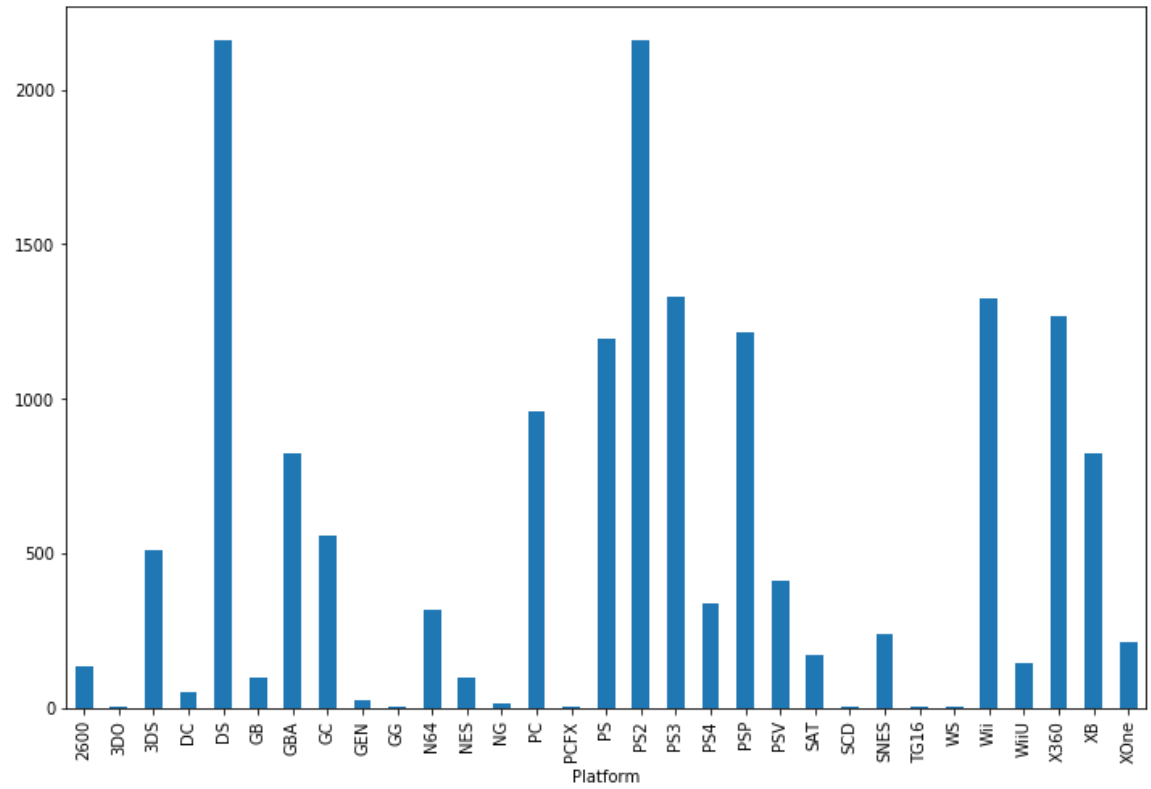
```
x = data.groupby(['Year']).count()
gamenum = x['Global_Sales']
plt.figure(figsize=(12,8))
gamenum.plot.bar()
plt.show()
#由图中可见，游戏的销售量与发行量存在正相关关系，销售量的变化趋势与发行量的变化趋势相似。
```



各发行平台的销售量比较

In [14]:

```
x = data.groupby(['Platform']).count()
gamenum = x['Global_Sales']
plt.figure(figsize=(12,8))
gamenum.plot.bar()
plt.show()
#由图中可见，DS平台和PS2平台发行量最高，平台之间的销售量差距较大，几个大型的发行平台销售亮比较可观，但是也存在小众的发行平台销售量与其他平台差距较大。
```



各发布人发布游戏销售量对比

In [15]:

```
x = data.groupby(['Publisher']).count()
gamenum = x[:30]['Global_Sales']
plt.figure(figsize=(12,8))
gamenum.plot.bar()
plt.show()
#由于Publisher数量众多，图中选取了30个作为可视化展示，由图中可见，各发布人之间发布游戏的销售量差距很大，只有少量发布人发布游戏销售量极为可观，而其他发布人大多销售量较少。
```