In [35]:

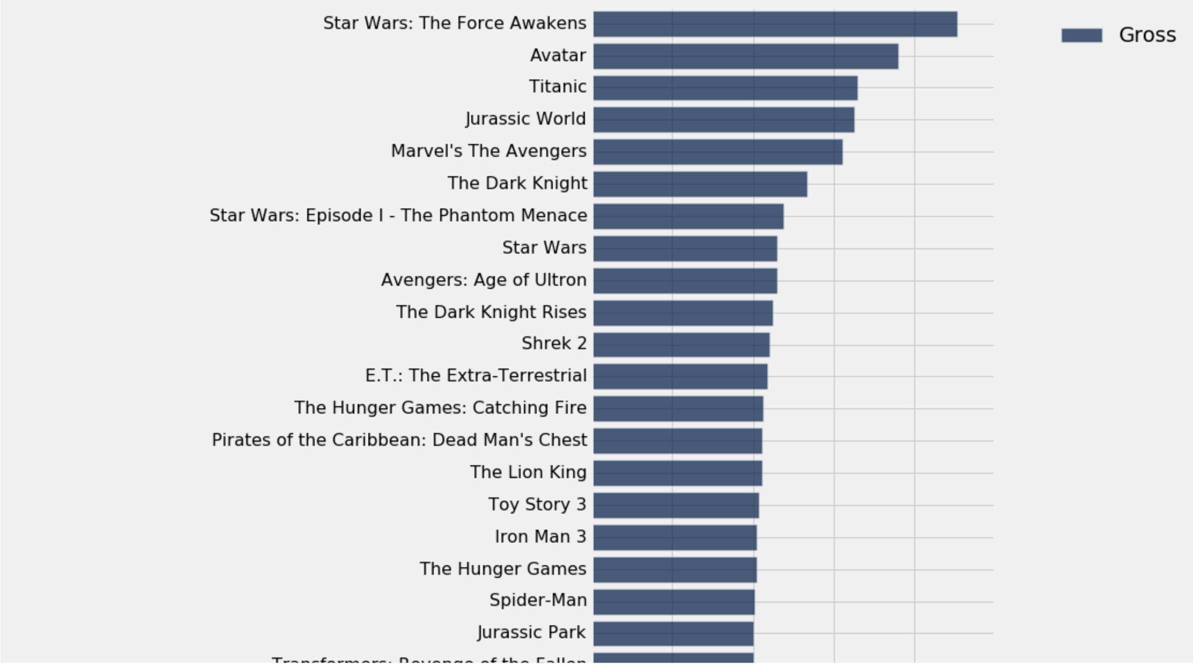
```
import numpy as np
from datascience import *

# Configure notebook (happens automatically on data8.berkeley.edu)
%matplotlib inline
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')

# Configure for presentation
np.set_printoptions(threshold=50, linewidth=50)
import matplotlib as mpl
mpl.rc('font', size=16)
```

In [125]:
```
top = Table.read_table('top_movies.csv')
top.set_format([2, 3], NumberFormatter).show(30)
```

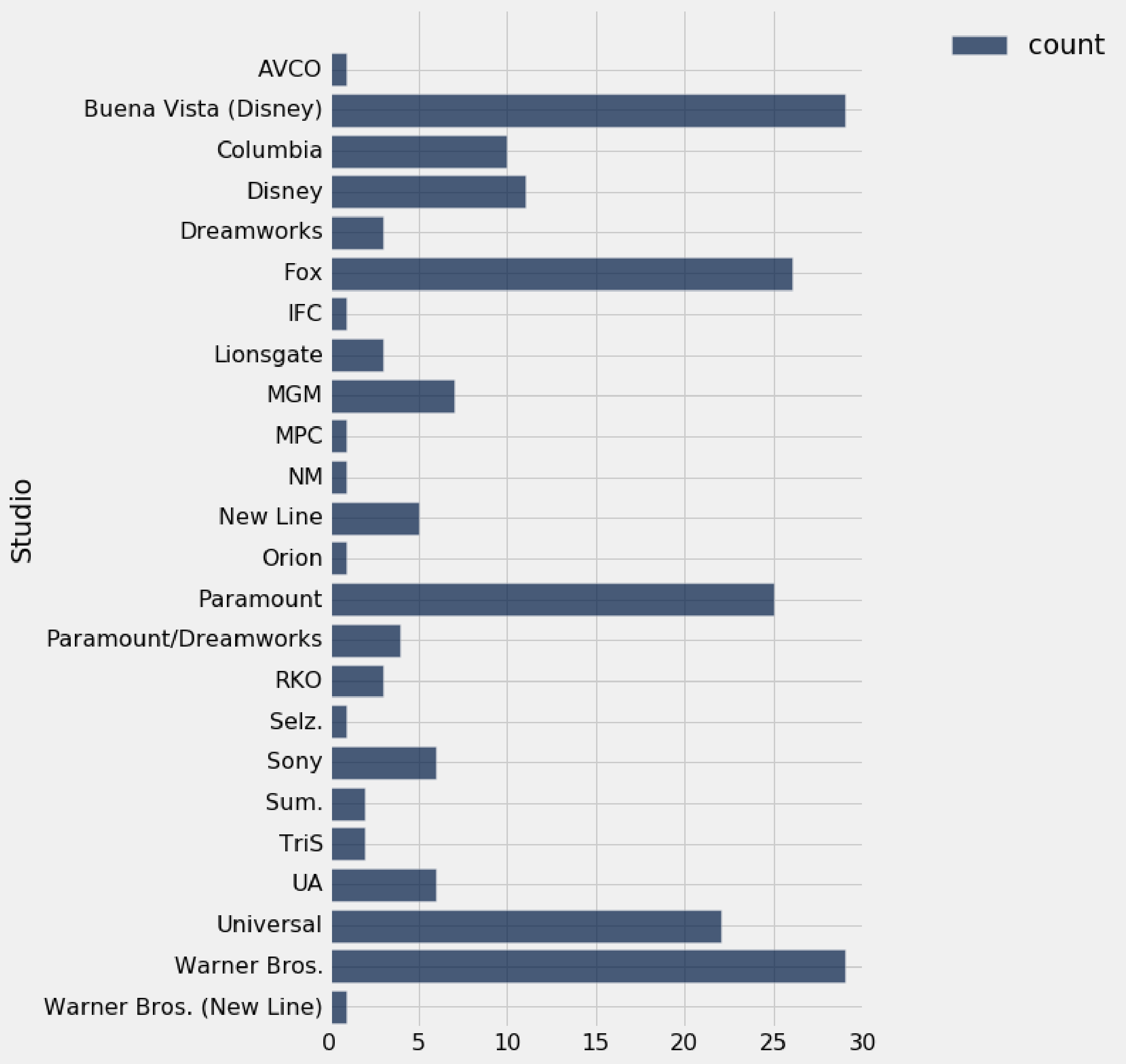
| Title | Studio | Gross | Gross (Adjusted) | Year |
|---:|---:|---:|---:|---:|
| Star Wars: The Force Awakens | Buena Vista (Disney) | 906,723,418 | 906,723,400 | 2015 |
| Avatar | Fox | 760,507,625 | 846,120,800 | 2009 |
| Titanic | Paramount | 658,672,302 | 1,178,627,900 | 1997 |
| Jurassic World | Universal | 652,270,625 | 687,728,000 | 2015 |
| Marvel's The Avengers | Buena Vista (Disney) | 623,357,910 | 668,866,600 | 2012 |
| The Dark Knight | Warner Bros. | 534,858,444 | 647,761,600 | 2008 |
| Star Wars: Episode I - The Phantom Menace | Fox | 474,544,677 | 785,715,000 | 1999 |
| Star Wars | Fox | 460,998,007 | 1,549,640,500 | 1977 |
| Avengers: Age of Ultron | Buena Vista (Disney) | 459,005,868 | 465,684,200 | 2015 |
| The Dark Knight Rises | Warner Bros. | 448,139,099 | 500,961,700 | 2012 |
| Shrek 2 | Dreamworks | 441,226,247 | 618,143,100 | 2004 |
| E.T.: The Extra-Terrestrial | Universal | 435,110,554 | 1,234,132,700 | 1982 |
| The Hunger Games: Catching Fire | Lionsgate | 424,668,047 | 444,697,400 | 2013 |
| Pirates of the Caribbean: Dead Man's Chest | Buena Vista (Disney) | 423,315,812 | 562,266,800 | 2006 |
| The Lion King | Buena Vista (Disney) | 422,783,777 | 775,573,900 | 1994 |
| Toy Story 3 | Buena Vista (Disney) | 415,004,880 | 454,156,300 | 2010 |
| Iron Man 3 | Buena Vista (Disney) | 409,013,994 | 424,632,700 | 2013 |
| The Hunger Games | Lionsgate | 408,010,692 | 442,510,400 | 2012 |
| Spider-Man | Sony | 403,706,375 | 604,517,300 | 2002 |
| Jurassic Park | Universal | 402,453,882 | 799,721,000 | 1993 |
| Transformers: Revenge of the Fallen | Paramount/Dreamworks | 402,111,870 | 468,938,100 | 2009 |
| Frozen | Buena Vista (Disney) | 400,738,009 | 426,656,900 | 2013 |
| Harry Potter and the Deathly Hallows Part 2 | Warner Bros. | 381,011,219 | 417,512,200 | 2011 |
| Finding Nemo | Buena Vista (Disney) | 380,843,261 | 535,802,700 | 2003 |
| Star Wars: Episode III - Revenge of the Sith | Fox | 380,270,577 | 516,123,900 | 2005 |
| The Lord of the Rings: The Return of the King | New Line | 377,845,905 | 536,265,400 | 2003 |
| Spider-Man 2 | Sony | 373,585,825 | 523,381,100 | 2004 |
| The Passion of the Christ | NM | 370,782,930 | 519,432,100 | 2004 |
| Despicable Me 2 | Universal | 368,061,265 | 407,978,700 | 2013 |
| Inside Out | Buena Vista (Disney) | 356,461,711 | 375,723,400 | 2015 |

... (170 rows omitted)

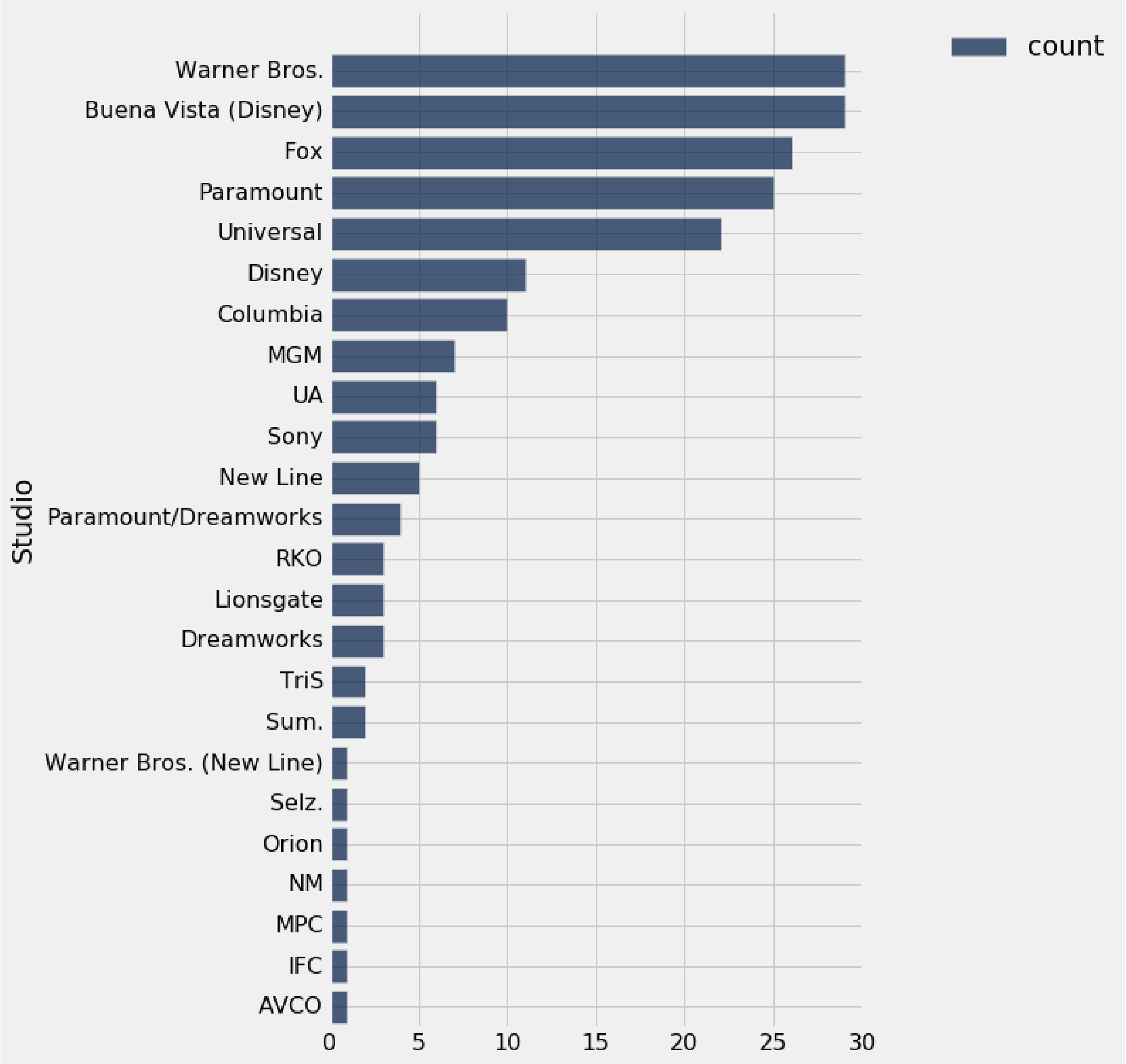
In [127]:
```
top.barh('Title', 'Gross')
```



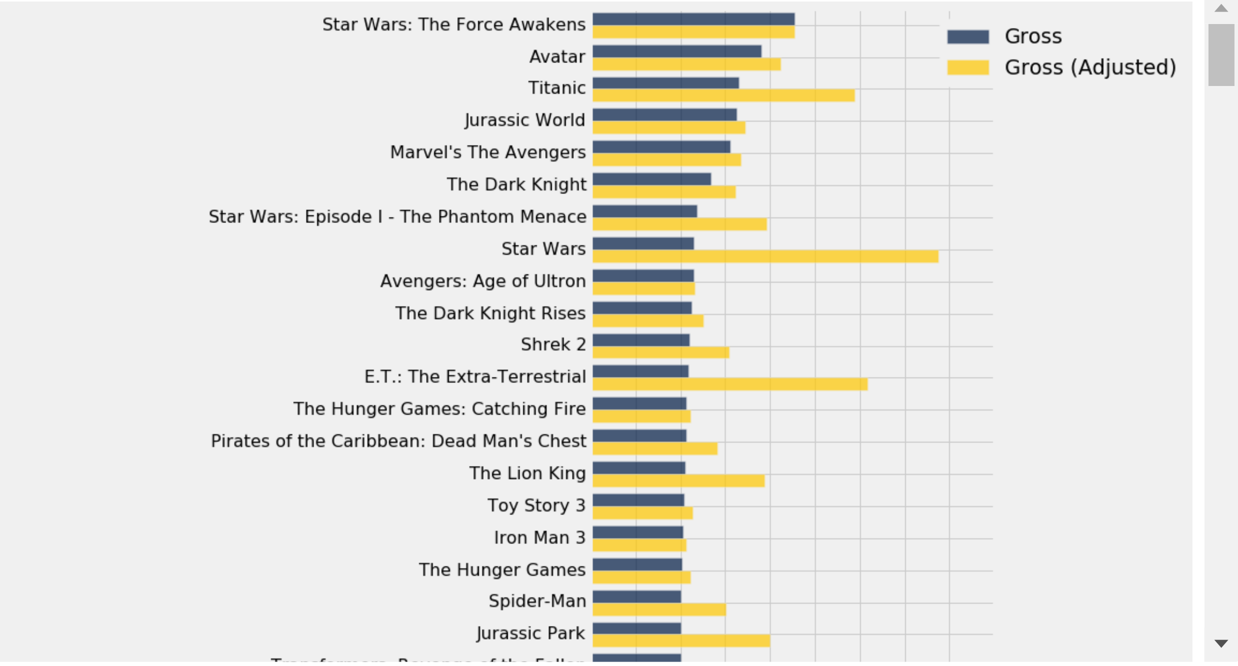
```
In [129]: top.group('Studio').barh('Studio', 'count')
```



```
In [130]: top.group('Studio').sort('count', descending=True).barh('Studio', 'count')
```



In [132]:
```
top.select([0, 2, 3]).barh('Title')
```



## Histograms

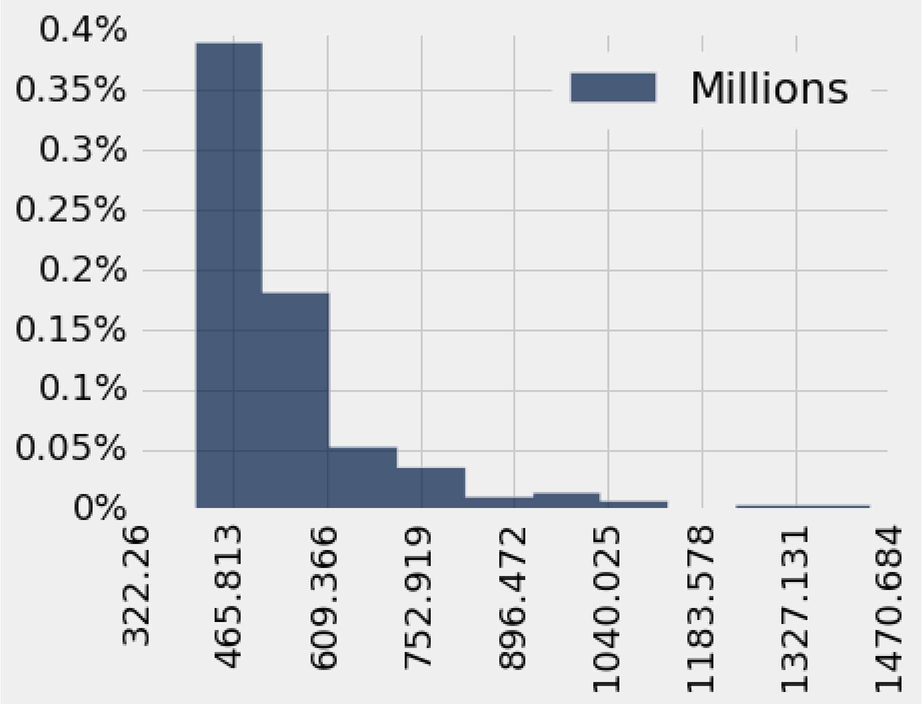
In [32]:
```
in_millions = np.round(top.column(3)/1000000, 2)
millions = top.select(['Title', 'Year']).with_column('Millions', in_millions)
millions
```

Out[32]:

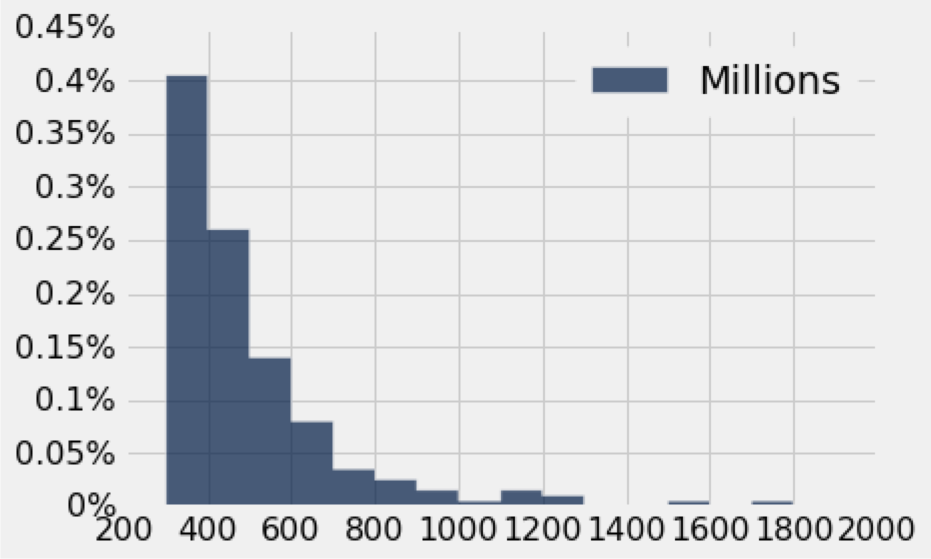
| Title | Year | Millions |
|---|---|---|
| Star Wars: The Force Awakens | 2015 | 906.72 |
| Avatar | 2009 | 846.12 |
| Titanic | 1997 | 1178.63 |
| Jurassic World | 2015 | 687.73 |
| Marvel's The Avengers | 2012 | 668.87 |
| The Dark Knight | 2008 | 647.76 |
| Star Wars: Episode I - The Phantom Menace | 1999 | 785.72 |
| Star Wars | 1977 | 1549.64 |
| Avengers: Age of Ultron | 2015 | 465.68 |
| The Dark Knight Rises | 2012 | 500.96 |

... (190 rows omitted)

```
In [33]: millions.hist('Millions')
```



```
In [45]: millions.hist('Millions', bins=np.arange(300,2001,100))
```



In [55]:
```
millions.bin('Millions', bins=np.arange(300,2001,100))
```

Out[55]:

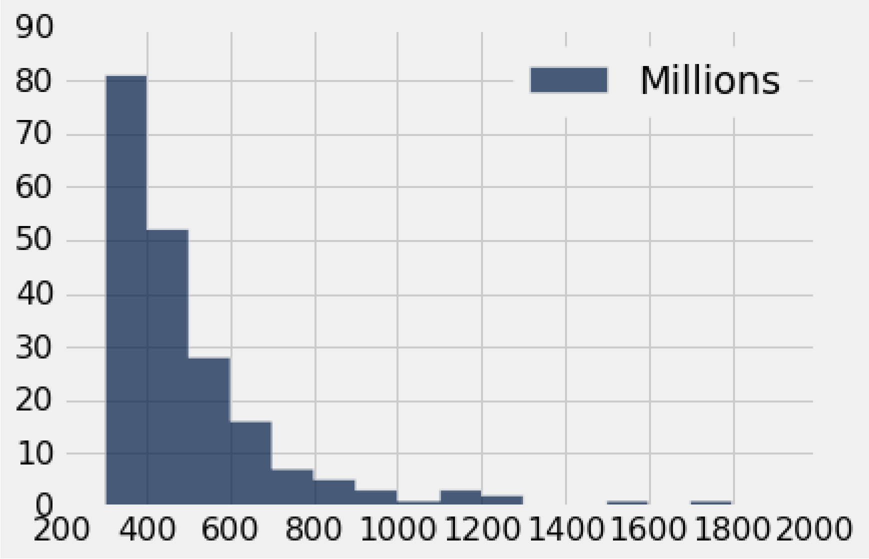
| bin | Millions count |
|---|---|
| 300 | 81 |
| 400 | 52 |
| 500 | 28 |
| 600 | 16 |
| 700 | 7 |
| 800 | 5 |
| 900 | 3 |
| 1000 | 1 |
| 1100 | 3 |
| 1200 | 2 |

... (8 rows omitted)

In [81]:
```
PercentFormatter().format_value(81/millions.num_rows/(400-300))
```

Out[81]:
```
'0.41%'
```

In [82]:
```
millions.hist('Millions', bins=np.arange(300,2001,100), normed=False)
```

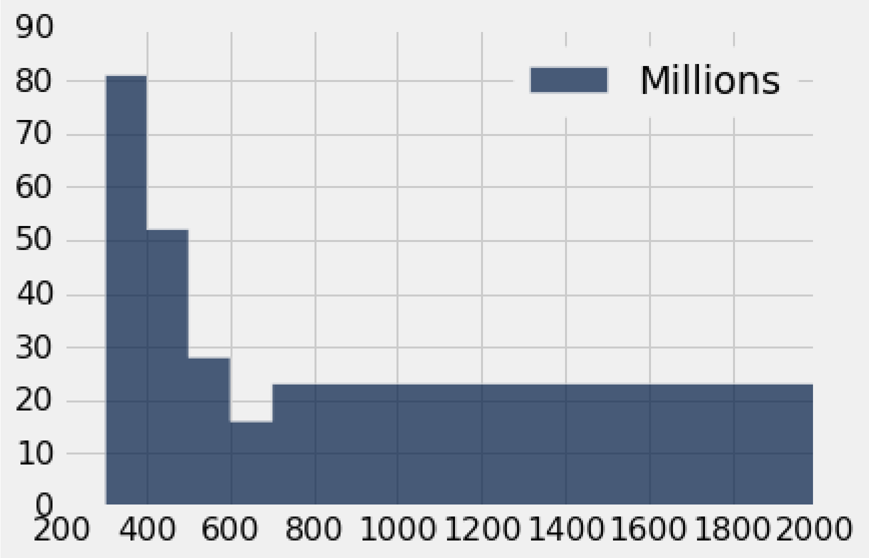


In [90]:
```
millions.bin('Millions', bins=[300, 400, 500, 600, 700, 2000])
```

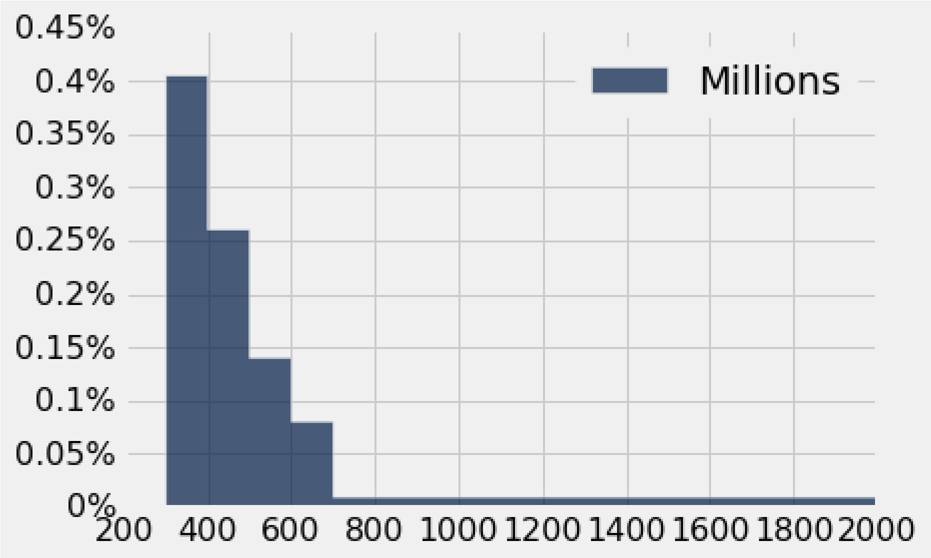
Out[90]:

| bin | Millions count |
|---|---|
| 300 | 81 |
| 400 | 52 |
| 500 | 28 |
| 600 | 16 |
| 700 | 23 |
| 2000 | 0 |

In [92]:
```
millions.hist('Millions', bins=[300, 400, 500, 600, 700, 2000], normed=False)
```



In [91]:
```
millions.hist('Millions', bins=[300, 400, 500, 600, 700, 2000])
```



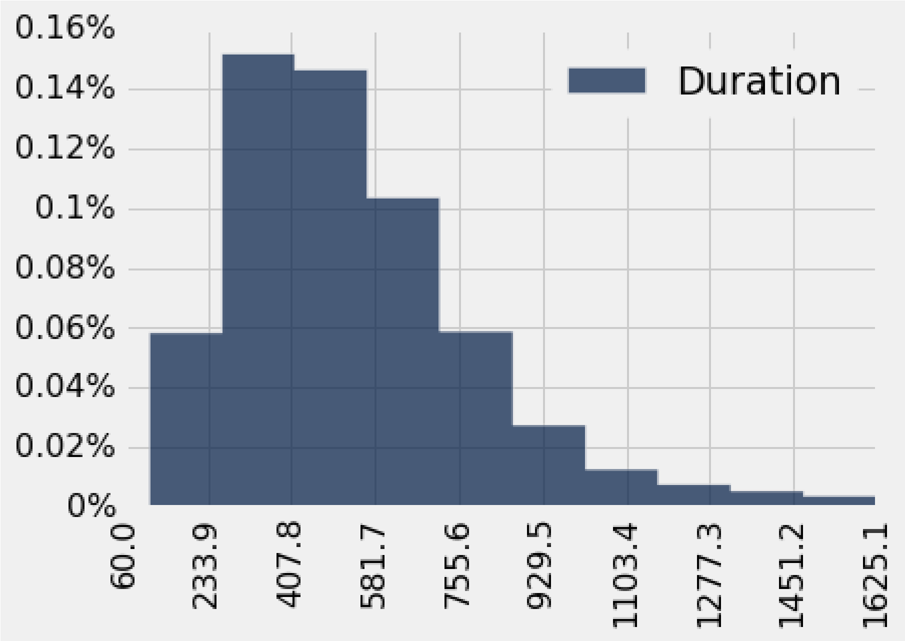
In [104]:
```
trips = Table.read_table('trip.csv')
trips
```

Out[104]:

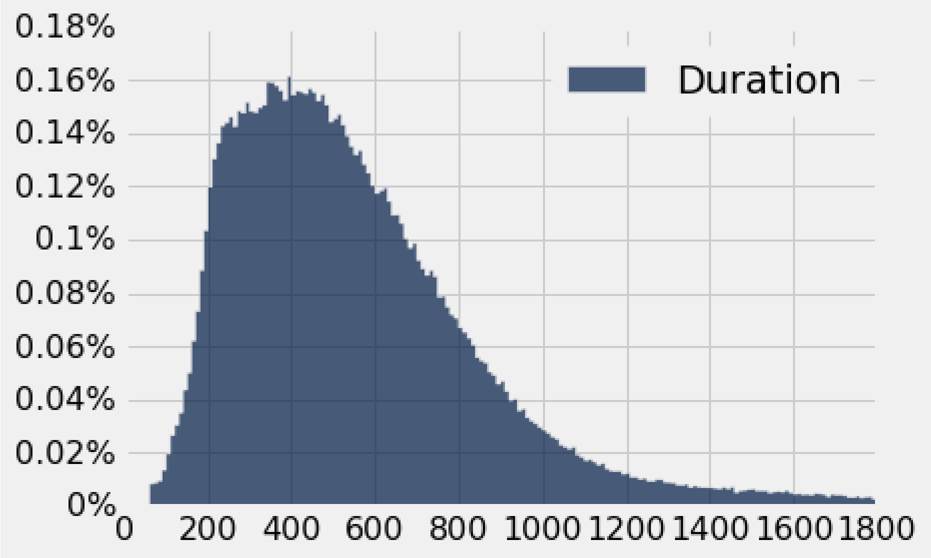
| Trip ID | Duration | Start Date | Start Station | Start Terminal | End Date | End Station | End Terminal | Bike # | Subsc |
|---|---|---|---|---|---|---|---|---|---|
| 913460 | 765 | 8/31/2015 23:26 | Harry Bridges Plaza (Ferry Building) | 50 | 8/31/2015 23:39 | San Francisco Caltrain (Townsend at 4th) | 70 | 288 | Subs |
| 913459 | 1036 | 8/31/2015 23:11 | San Antonio Shopping Center | 31 | 8/31/2015 23:28 | Mountain View City Hall | 27 | 35 | Subs |
| 913455 | 307 | 8/31/2015 23:13 | Post at Kearny | 47 | 8/31/2015 23:18 | 2nd at South Park | 64 | 468 | Subs |
| 913454 | 409 | 8/31/2015 23:10 | San Jose City Hall | 10 | 8/31/2015 23:17 | San Salvador at 1st | 8 | 68 | Subs |
| 913453 | 789 | 8/31/2015 23:09 | Embarcadero at Folsom | 51 | 8/31/2015 23:22 | Embarcadero at Sansome | 60 | 487 | Cust |
| 913452 | 293 | 8/31/2015 23:07 | Yerba Buena Center of the Arts (3rd @ Howard) | 68 | 8/31/2015 23:12 | San Francisco Caltrain (Townsend at 4th) | 70 | 538 | Subs |
| 913451 | 896 | 8/31/2015 23:07 | Embarcadero at Folsom | 51 | 8/31/2015 23:22 | Embarcadero at Sansome | 60 | 363 | Cust |
| 913450 | 255 | 8/31/2015 22:16 | Embarcadero at Sansome | 60 | 8/31/2015 22:20 | Steuart at Market | 74 | 470 | Subs |
| 913449 | 126 | 8/31/2015 22:12 | Beale at Market | 56 | 8/31/2015 22:15 | Temporary Transbay Terminal (Howard at Beale) | 55 | 439 | Subs |
| 913448 | 932 | 8/31/2015 21:57 | Post at Kearny | 47 | 8/31/2015 22:12 | South Van Ness at Market | 66 | 472 | Subs |

... (354142 rows omitted)

In [124]:
```
commute = trips.where(trips.column('Duration') < 1800)
commute.hist('Duration')
```



In [123]:
```
commute.hist('Duration', bins=np.arange(60, 1801, 10))
```


Duration