In [5]:

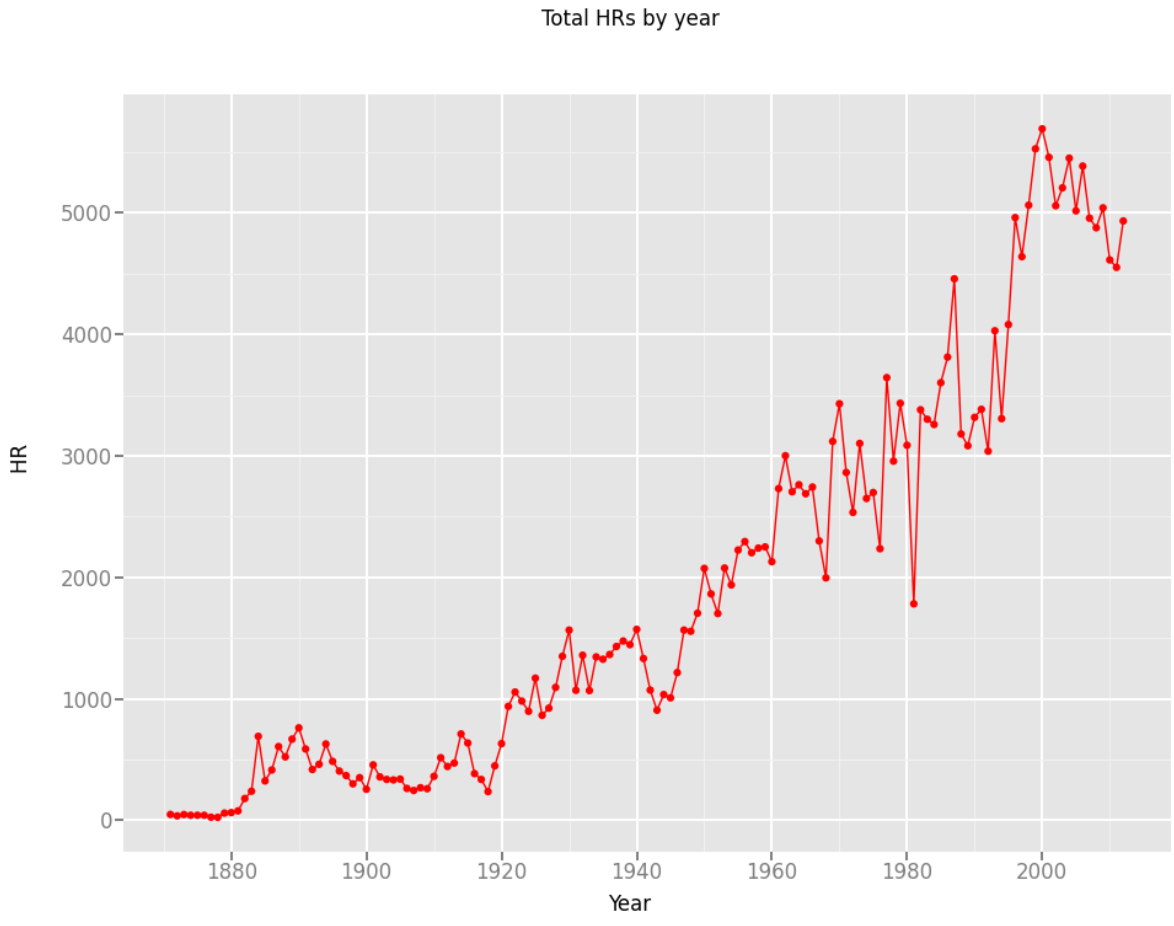
```
from pandas import *
from ggplot import *

import pandas

def lineplot(hr_year_csv):
    # A csv file will be passed in as an argument which
    # contains two columns -- 'HR' (the number of homerun hits)
    # and 'yearID' (the year in which the homeruns were hit).
    #
    # Fill out the body of this function, lineplot, to use the
    # passed-in csv file, hr_year.csv, and create a
    # chart with points connected by lines, both colored 'red',
    # showing the number of HR by year.
    #
    # You will want to first load the csv file into a pandas dataframe
    # and use the pandas dataframe along with ggplot to create your visualization
    #
    # You can check out the data in the csv file at the link below:
    # https://s3.amazonaws.com/content.udacity-data.com/courses/ud359/hr_year.csv
    #
    # You can read more about ggplot at the following link:
    # https://github.com/yhat/ggplot/
    hr_year = pandas.read_csv(hr_year_csv)
    gg = ggplot(hr_year, aes('yearID', 'HR')) + \
        geom_point(color='red') + \
        geom_line(color='red') + \
        ggtitle('Total HRs by year') + \
        xlab('Year') + \
        ylab('HR ')
    return gg

if __name__ == "__main__":

    print(lineplot('hr_year.csv'))
```




```
<ggplot: (-9223371932862393228)>
```

In [7]:

```
import pandas
from ggplot import *

def lineplot_compare(hr_by_team_year_sf_la_csv):
    # Write a function, lineplot_compare, that will read a csv file
    # called hr_by_team_year_sf_la.csv and plot it using pandas and ggplot.
    #
    # This csv file has three columns: yearID, HR, and teamID. The data in the
    # file gives the total number of home runs hit each year by the SF Giants
    # (teamID == 'SFN') and the LA Dodgers (teamID == "LAN"). Produce a
    # visualization comparing the total home runs by year of the two teams.
    #
    # You can see the data in hr_by_team_year_sf_la_csv
    # at the link below:
    # https://s3.amazonaws.com/content.udacity-data.com/courses/ud359/hr_by_team_year_sf_la.csv
    #
    # Note that to differentiate between multiple categories on the
    # same plot in ggplot, we can pass color in with the other arguments
    # to aes, rather than in our geometry functions. For example,
    # ggplot(data, aes(xvar, yvar, color=category_var)). This might help you
    # in this exercise.
    data = read_csv(hr_by_team_year_sf_la_csv)
    df = DataFrame(data)
    gg = ggplot(df, aes('HR','yearID',color='teamID')) + geom_point() + geom_line() + ggtitle('Number of HR by year') + xlab('HR'

    return gg


if __name__ == "__main__":

    print(lineplot('hr_year.csv'))
```