In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_csv("files_relevant/all_matches.tsv", sep='\t')
df
```

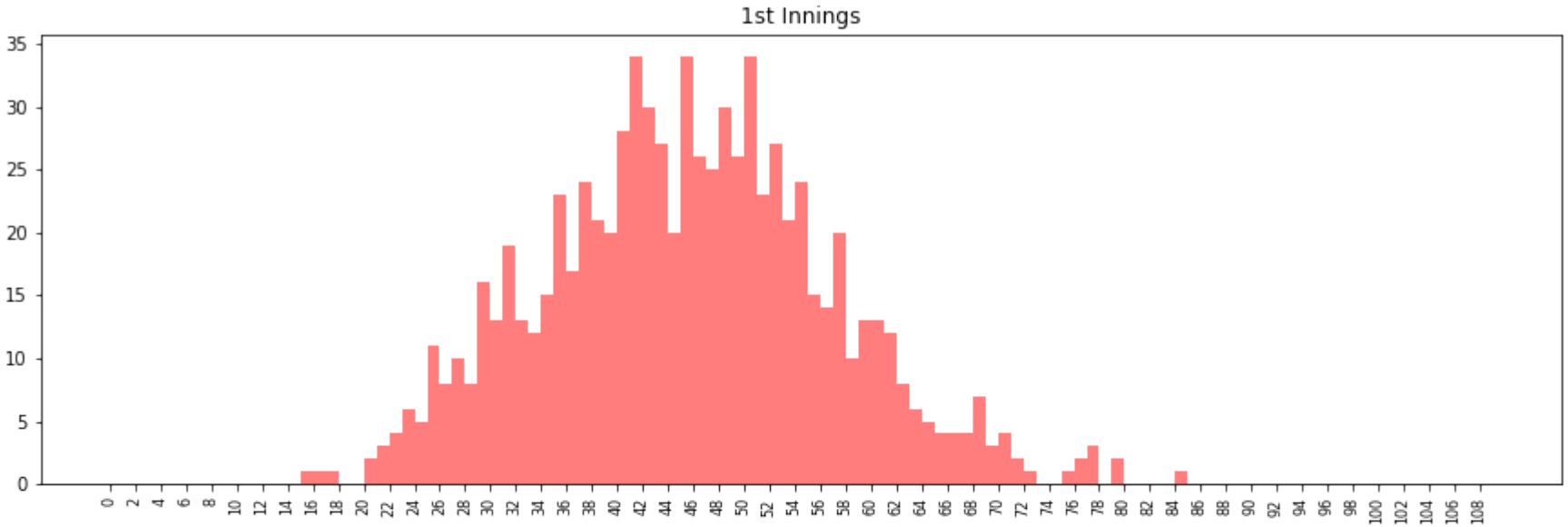
Out[2]:

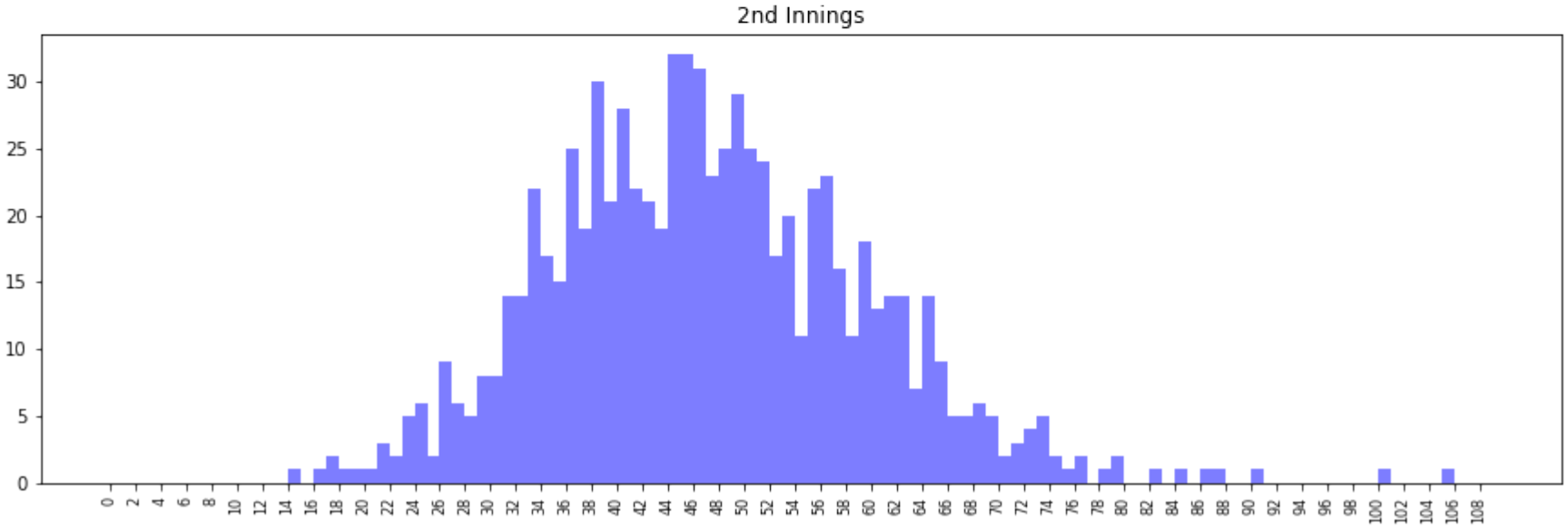
	Unnamed: 0	venue	innings	ball	batting_team	bowling_team	batsmen	bowlers	runs
0	0	M Chinnaswamy Stadium	1	5.6	Kolkata Knight Riders	Royal Challengers Bangalore	RT Ponting,BB McCullum,SC Ganguly	Z Khan,AA Noffke,P Kumar	61
1	1	M Chinnaswamy Stadium	2	5.6	Royal Challengers Bangalore	Kolkata Knight Riders	MV Boucher,V Kohli,R Dravid,W Jaffer,JH Kallis...	AB Agarkar,AB Dinda,I Sharma	26
2	10	Sawai Mansingh Stadium	1	5.6	Kings XI Punjab	Rajasthan Royals	K Goel,JR Hopes,KC Sangakkara	MM Patel,SR Watson,SK Trivedi	54
3	100	Eden Gardens	2	5.6	Kolkata Knight Riders	Kings XI Punjab	SC Ganguly,Salman Butt,DB Das,DJ Hussey	IK Pathan,RR Powar,S Sreesanth	51
4	1000	Rajiv Gandhi International Stadium	1	5.6	Sunrisers Hyderabad	Kings XI Punjab	S Dhawan,DA Warner	BE Hendricks,Sandeep Sharma,Gurkeerat Singh,An...	56
...	...	...	...	...	...	...	...	...	...
1619	995	Shaheed Veer Narayan Singh International Stadium	2	5.6	Delhi Daredevils	Sunrisers Hyderabad	JP Duminy,Q de Kock,SS Iyer	Parvez Rasool,P Kumar,B Kumar,I Sharma	49
1620	996	Wankhede Stadium	1	5.6	Royal Challengers Bangalore	Mumbai Indians	CH Gayle,V Kohli,AB de Villiers	JJ Bumrah,MJ McClenaghan,J Suchith,SL Malinga	39
1621	997	Wankhede Stadium	2	5.6	Mumbai Indians	Royal Challengers Bangalore	PA Patel,LMP Simmons,RG Sharma	S Aravind,MA Starc,D Wiese,HV Patel	45
1622	998	MA Chidambaram Stadium	1	5.6	Chennai Super Kings	Rajasthan Royals	DR Smith,SK Raina,BB McCullum,F du Plessis	PV Tambe,CH Morris,SR Watson,Ankit Sharma	36
1623	999	MA Chidambaram Stadium	2	5.6	Rajasthan Royals	Chennai Super Kings	SR Watson,SPD Smith,AM Rahane	A Nehra,MM Sharma	40

1624 rows × 9 columns

In [3]:
```
plt.rcParams['xtick.labelsize'] = 8

df_innings_1 = df.loc[df["innings"] == 1]
```

```
df_innings_2 = df.loc[df["innings"] == 2]

plt.figure(figsize=(15,10))

plt.subplot(211)
plt.xticks(rotation=90, ticks = [i for i in range(0, 110, 2)])
plt.title("1st Innings")
plt.hist(df_innings_1["runs"], bins = [x for x in range(110)], color='r', alpha = 0.5)

plt.subplot(212)
plt.xticks(rotation=90, ticks = [i for i in range(0, 110, 2)])
plt.title("2nd Innings")
plt.hist(df_innings_2["runs"], bins = [x for x in range(110)], color='b', alpha = 0.5)

plt.show()
```





In [4]:
```
df_innings_1.describe()
```

Out[4]:

	Unnamed: 0	innings	ball	runs
count	816.000000	816.0	816.000000	816.000000
mean	810.919118	1.0	5.620343	44.986520
std	468.862624	0.0	0.047026	11.359537
min	0.000000	1.0	5.600000	15.000000
25%	405.500000	1.0	5.600000	37.000000
50%	813.000000	1.0	5.600000	45.000000
75%	1215.500000	1.0	5.600000	52.000000
max	1622.000000	1.0	5.900000	84.000000

In [5]:
```
df_innings_2.describe()
```

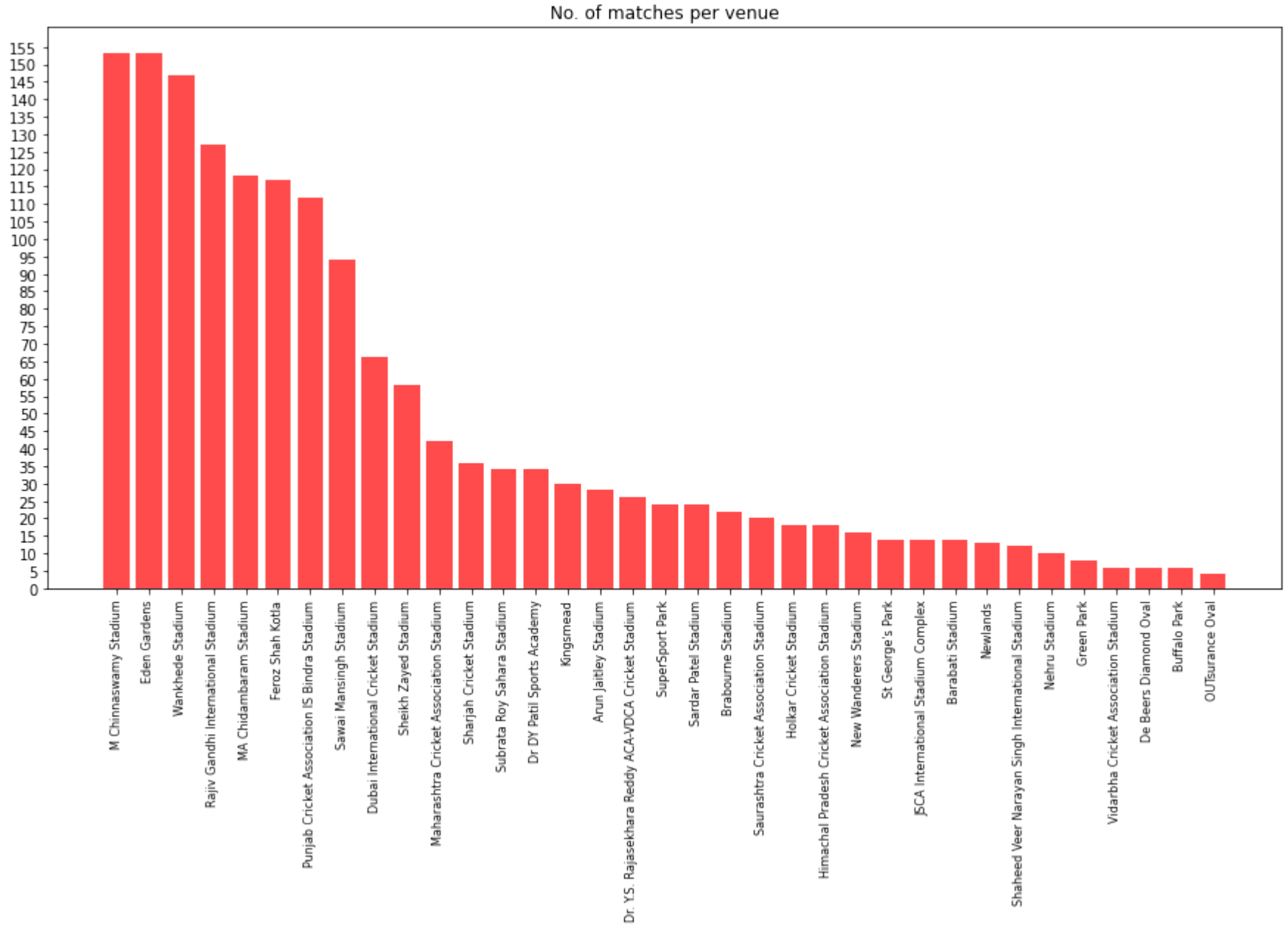
Out[5]:

	Unnamed: 0	innings	ball	runs
count	808.000000	808.0	808.000000	808.000000
mean	812.086634	2.0	5.619554	46.695545
std	469.333396	0.0	0.046314	12.396041
min	1.000000	2.0	5.600000	14.000000
25%	406.500000	2.0	5.600000	38.000000
50%	810.000000	2.0	5.600000	46.000000
75%	1218.500000	2.0	5.600000	55.000000
max	1623.000000	2.0	5.900000	105.000000

In [6]:
```
unique_venue = df.venue.unique()
count = []

for venue in unique_venue:
    count.append(df.loc[df["venue"]==venue, "venue"].count())
```

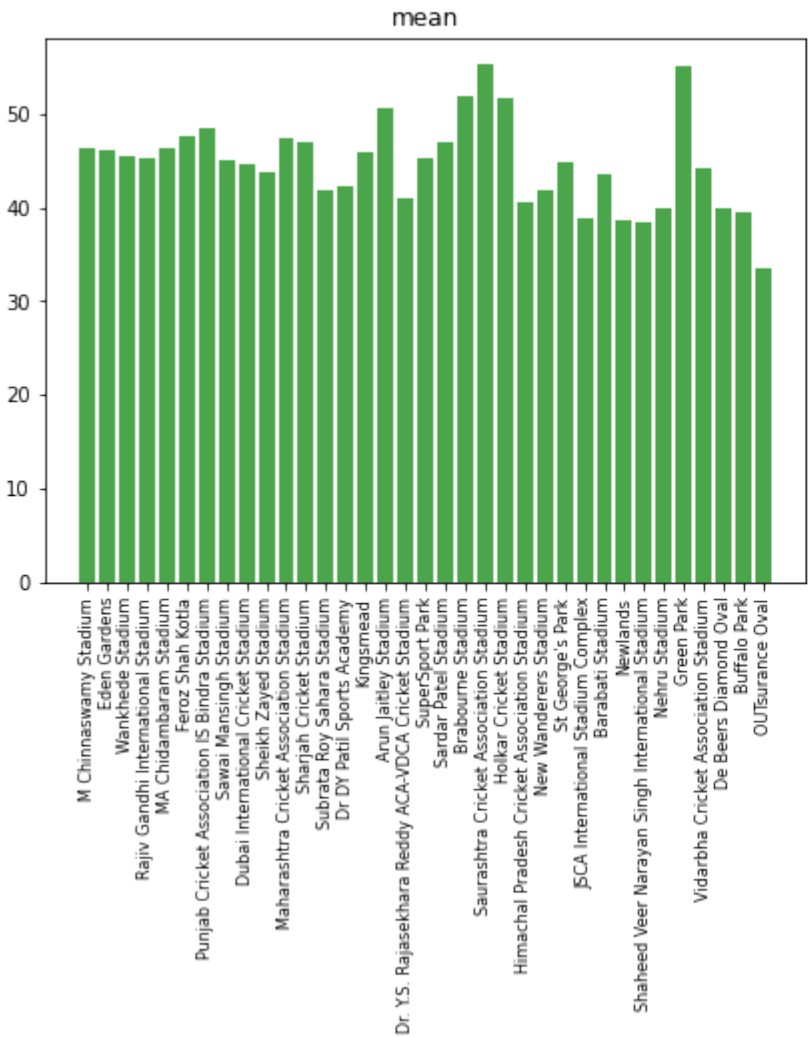
```
unique_venue = sorted(list(zip(count, unique_venue)), reverse = True)
count = list(zip(*unique_venue))[0]
unique_venue = list(zip(*unique_venue))[1]

print(f"Number of stadiums: {len(unique_venue)}")

plt.figure(figsize=(15, 7))

plt.xticks(rotation=90)
plt.yticks(ticks=[x for x in range(0, 160, 5)])
plt.title("No. of matches per venue")
plt.bar(x = unique_venue, height = count, color='r', alpha=0.7)
plt.show()
```

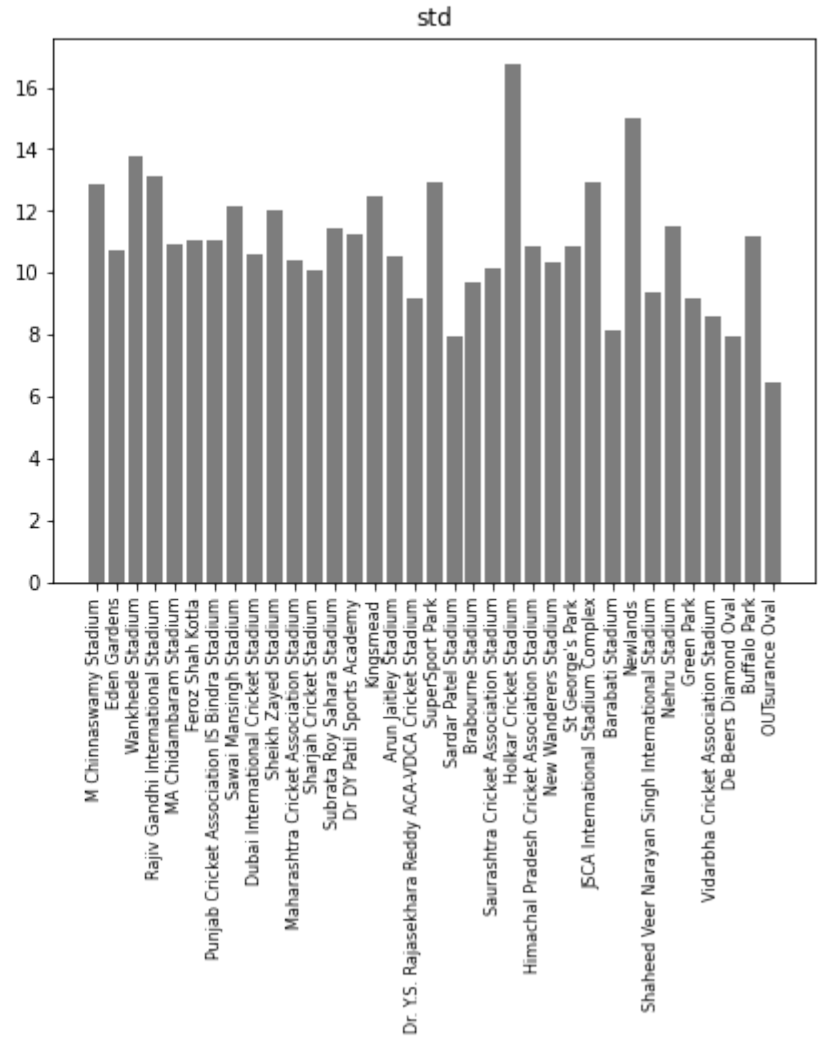
```
Number of stadiums: 35
```


No. of matches per venue
155
150
145
140
135
130
125
120
115
110
105
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
M Chinnaswamy Stadium
Eden Gardens
Wankhede Stadium
Rajiv Gandhi International Stadium
MA Chidambaram Stadium
Feroz Shah Kotla
Punjab Cricket Association IS Bindra Stadium
Sawai Mansingh Stadium
Dubai International Cricket Stadium
Sheikh Zayed Stadium
Maharashtra Cricket Association Stadium
Sharjah Cricket Stadium
Subrata Roy Sahara Stadium
Dr DY Patil Sports Academy
Kingsmead
Arun Jaitley Stadium
Dr. YS. Rajasekhara Reddy ACA-VDCA Cricket Stadium
SuperSport Park
Sardar Patel Stadium
Brabourne Stadium
Saurashtra Cricket Association Stadium
Holkar Cricket Stadium
Himachal Pradesh Cricket Association Stadium
New Wanderers Stadium
St George's Park
JSCA International Stadium Complex
Barabati Stadium
Newlands
Shaheed Veer Narayan Singh International Stadium
Nehru Stadium
Green Park
Vidarbha Cricket Association Stadium
De Beers Diamond Oval
Buffalo Park
OUTsurance Oval

In [7]:
```
mean = []
std = []
```

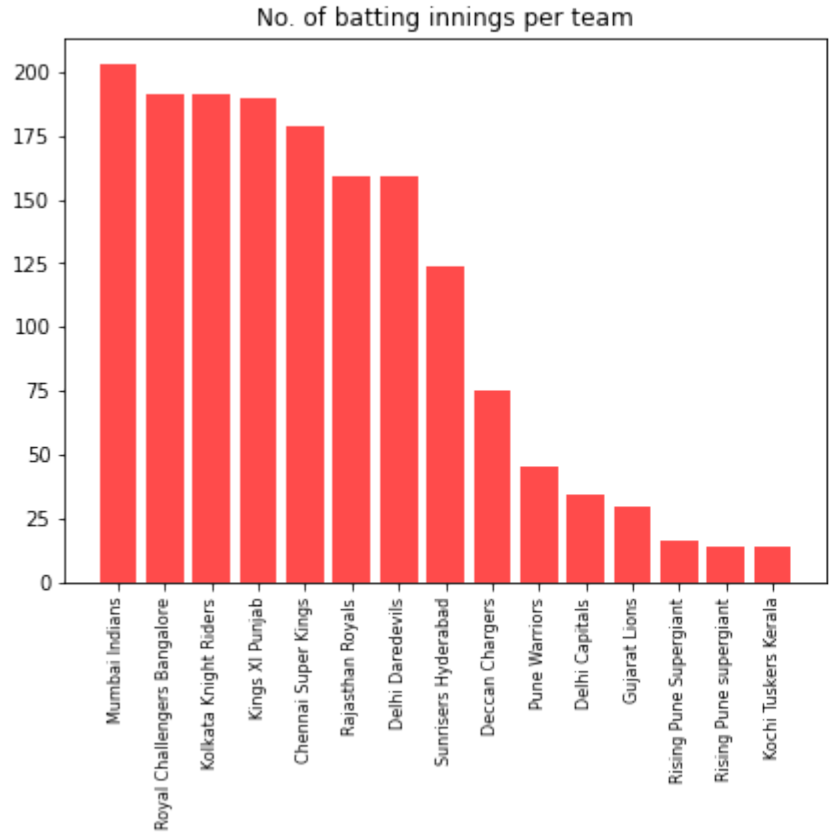
```
for venue in unique_venue:
    mean.append(df.loc[df["venue"]==venue, "runs"].mean())
    std.append(df.loc[df["venue"]==venue, "runs"].std())


plt.figure(figsize=(15, 5))

plt.subplot(121)
plt.xticks(rotation=90)
plt.title("mean")
plt.bar(x=unique_venue, height = mean, color='g', alpha=0.7)

plt.subplot(122)
plt.xticks(rotation=90)
plt.title('std')
plt.bar(x=unique_venue, height = std, color='black', alpha=0.5)

plt.show()
```

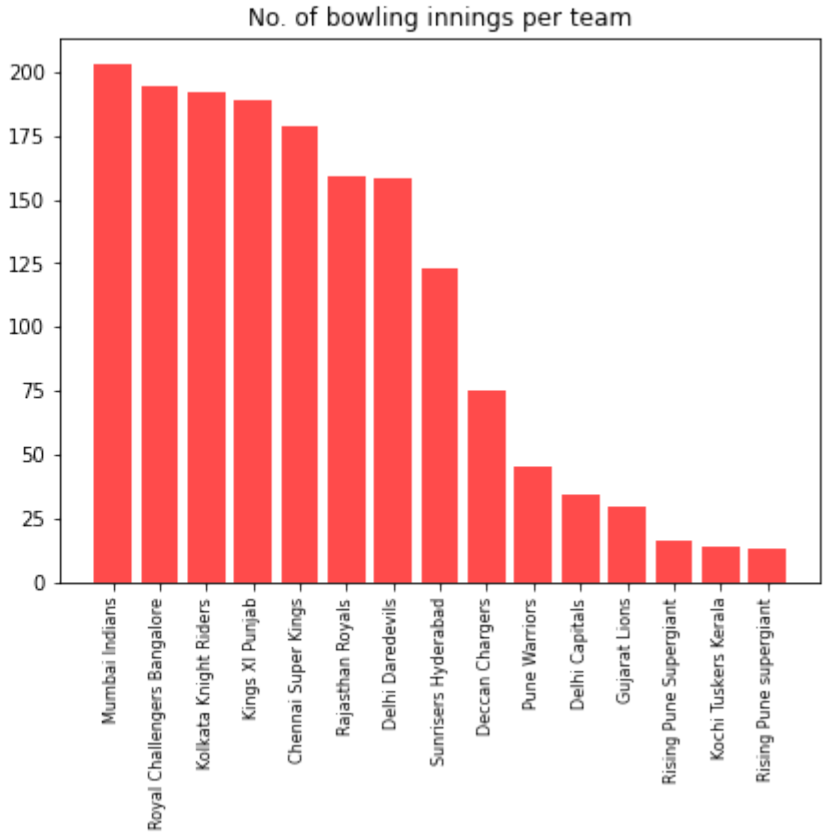




In [8]:
```
batting_teams = df.batting_team.unique()
bowling_teams = df.bowling_team.unique()

count_batting = []
count_bowling = []

for team in batting_teams:
    count_batting.append(df.loc[df["batting_team"]==team, "batting_team"].count())
```

```
for team in bowling_teams:
    count_bowling.append(df.loc[df["bowling_team"]==team, "bowling_team"].count())

batting_teams = sorted(list(zip(count_batting, batting_teams)), reverse = True)
count_batting = list(zip(*batting_teams))[0]
batting_teams = list(zip(*batting_teams))[1]

bowling_teams = sorted(list(zip(count_bowling, bowling_teams)), reverse = True)
count_bowling = list(zip(*bowling_teams))[0]
bowling_teams = list(zip(*bowling_teams))[1]

plt.figure(figsize=(15, 5))

plt.subplot(121)
plt.xticks(rotation=90)
# plt.yticks(ticks=[x for x in range(0, 160, 5)])
plt.title("No. of batting innings per team")
plt.bar(x = batting_teams, height = count_batting, color='r', alpha=0.7)

plt.subplot(122)
plt.xticks(rotation=90)
# plt.yticks(ticks=[x for x in range(0, 160, 5)])
plt.title("No. of bowling innings per team")
plt.bar(x = bowling_teams, height = count_bowling, color='r', alpha=0.7)

plt.show()
```

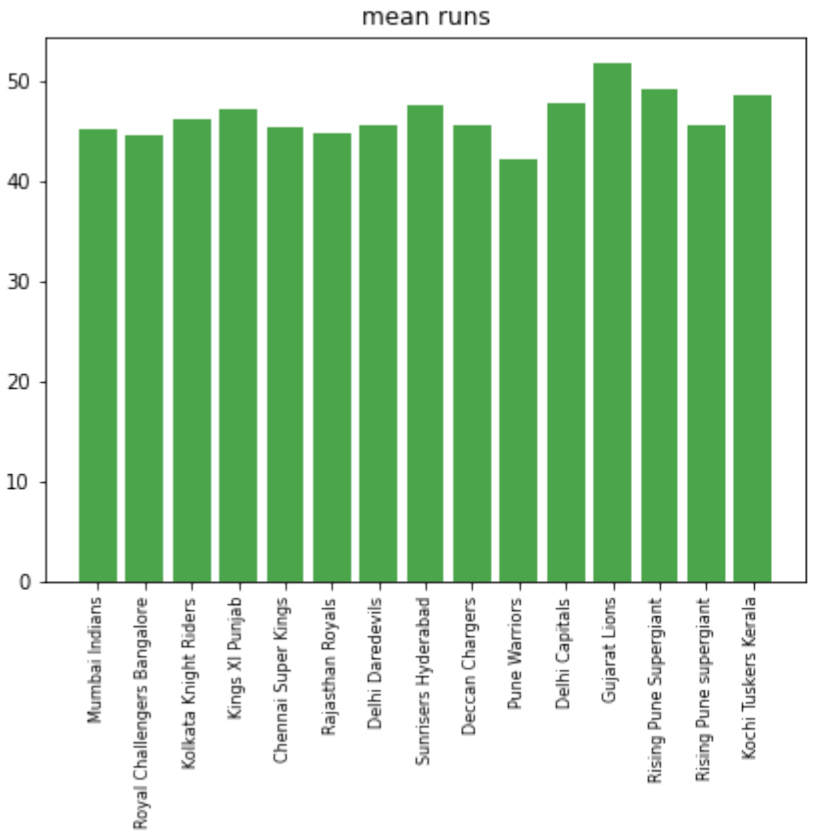


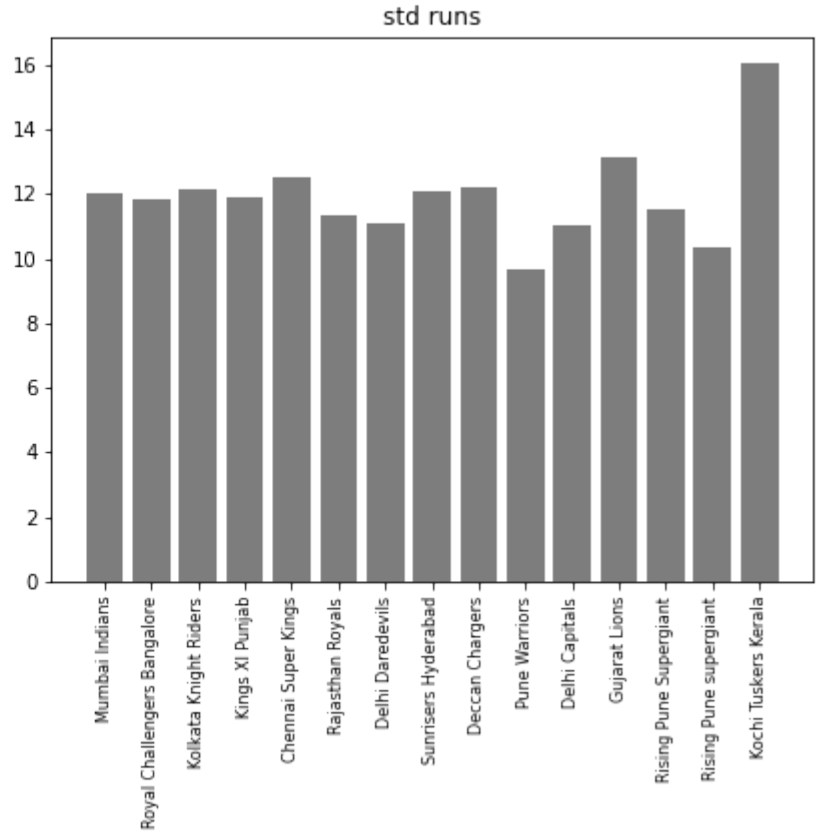


In [9]:
```
print(len(batting_teams))

mean_runs = []
std_runs = []

for team in batting_teams:
    mean_runs.append(df.loc[df["batting_team"]==team, "runs"].mean())
    std_runs.append(df.loc[df["batting_team"]==team, "runs"].std())

plt.figure(figsize=(15, 5))

plt.subplot(121)
plt.xticks(rotation=90)
plt.title("mean runs")
```

```
plt.bar(x=batting_teams, height = mean_runs, color='g', alpha=0.7)

plt.subplot(122)
plt.xticks(rotation=90)
plt.title('std runs')
plt.bar(x=batting_teams, height = std_runs, color='black', alpha=0.5)

plt.show()
```

15

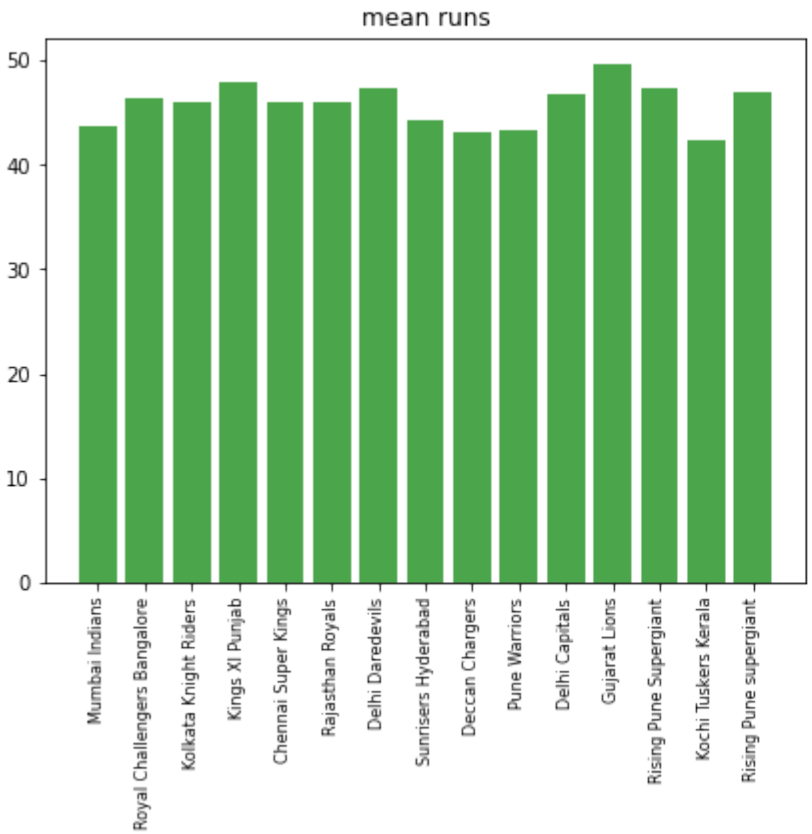


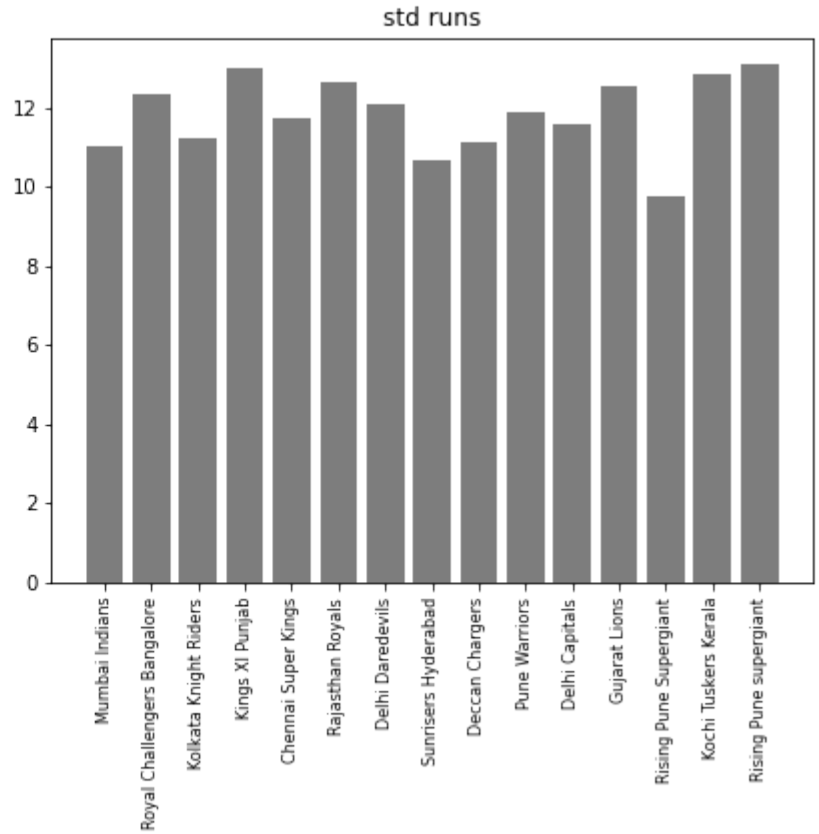


In [10]:
```
print(len(bowling_teams))

mean_runs = []
std_runs = []
```

```
for team in bowling_teams:
    mean_runs.append(df.loc[df["bowling_team"]==team, "runs"].mean())
    std_runs.append(df.loc[df["bowling_team"]==team, "runs"].std())

plt.figure(figsize=(15, 5))

plt.subplot(121)
plt.xticks(rotation=90)
plt.title("mean runs")
plt.bar(x=bowling_teams, height = mean_runs, color='g', alpha=0.7)

plt.subplot(122)
plt.xticks(rotation=90)
plt.title('std runs')
plt.bar(x=bowling_teams, height = std_runs, color='black', alpha=0.5)

plt.show()
```

15





In [11]:
```
unique_venue
```

Out[11]:
```
('M Chinnaswamy Stadium',
 'Eden Gardens',
 'Wankhede Stadium',
 'Rajiv Gandhi International Stadium',
 'MA Chidambaram Stadium',
 'Feroz Shah Kotla',
 'Punjab Cricket Association IS Bindra Stadium',
 'Sawai Mansingh Stadium',
 'Dubai International Cricket Stadium',
 'Sheikh Zayed Stadium',
 'Maharashtra Cricket Association Stadium',
 'Sharjah Cricket Stadium',
 'Subrata Roy Sahara Stadium',
 'Dr DY Patil Sports Academy',
 'Kingsmead',
 'Arun Jaitley Stadium',
 'Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium',
 'SuperSport Park',
 'Sardar Patel Stadium',
 'Brabourne Stadium',
 'Saurashtra Cricket Association Stadium',
 'Holkar Cricket Stadium',
 'Himachal Pradesh Cricket Association Stadium',
 'New Wanderers Stadium',
 "St George's Park",
 'JSCA International Stadium Complex',
 'Barabati Stadium',
 'Newlands',
 'Shaheed Veer Narayan Singh International Stadium',
 'Nehru Stadium',
 'Green Park',
 'Vidarbha Cricket Association Stadium',
 'De Beers Diamond Oval',
 'Buffalo Park',
 'OUTsurance Oval')
```