NQ

Gender: [χ^2:53.55, dof: 2, *p*-value: 2.4e-12]
- female: 82.14
- average: 71.38
- male: 70.98
- female_male: 23.68

Nationality: [χ^2: 5.01, dof: 5, *p*-value: 0.42]
- canada: 89.47
- uk: 74.80
- india: 73.17
- average: 71.38
- us: 71.04
- not_found: 68.42
- others: 67.74

Profession: [χ^2:22.17, dof:10, *p*-value: 0.014]
- film/tv: 77.56
- music: 76.15
- artist: 73.33
- average: 71.38
- fictional: 66.67
- film/tv_music: 60.98
- science/tech: 56.67
- not_found: 40.00

QuizBowl

Gender: [χ^2: 4.65, dof: 2, *p*-value: 0.098]
- female: 67.33
- male: 66.75
- average: 66.22
- female_male: 45.83

Nationality: [χ^2:17.61, dof:12, *p*-value: 0.13]
- ireland: 87.50
- spain: 75.00
- russia: 73.53
- germany: 73.33
- average: 66.22
- us: 62.82
- not_found: 59.30
- italy: 57.14
- netherlands: 55.56

Profession: [χ^2:48.20, dof: 9, *p*-value: 2.3e-07]
- writing: 76.76
- music: 71.79
- artist: 71.13
- average: 66.22
- not_found: 56.00
- science/tech: 50.75
- religion: 47.06
- politics_religion: 40.00

SQuAD

Gender: [χ^2: 0.91, dof: 1, *p*-value: 0.34]
- female: 92.73
- average: 87.67
- male: 87.45

Nationality: [χ^2:11.20, dof:14, *p*-value: 0.67]
- spain: 100.00
- india: 93.75
- egypt: 91.67
- uk: 91.48
- average: 87.67
- france: 84.47
- not_found: 82.76
- russia: 81.25
- greece: 77.78

Profession: [χ^2: 9.81, dof:11, *p*-value: 0.55]
- sports: 100.00
- religion: 96.30
- artist: 94.44
- economics: 89.29
- average: 87.67
- politics: 86.32
- music: 85.71
- philosophy: 82.61
- writing: 82.61

Trivia QA

Gender: [χ^2: 0.79, dof: 2, *p*-value: 0.67]
- female: 52.74
- female_male: 51.89
- average: 51.27
- male: 50.99

Nationality: [χ^2: 7.40, dof: 8, *p*-value: 0.49]
- others: 56.09
- uk_us: 55.05
- france: 53.66
- average: 51.27
- italy: 50.43
- uk: 50.15
- australia: 49.12
- germany: 48.39

Profession: [χ^2:27.95, dof:13, *p*-value: 0.0092]
- law/crime: 64.29
- science/tech: 58.58
- writing: 55.75
- film/tv_music: 53.38
- average: 51.27
- others: 50.52
- film/tv_writing: 49.09
- not_found: 48.28
- sports: 41.73

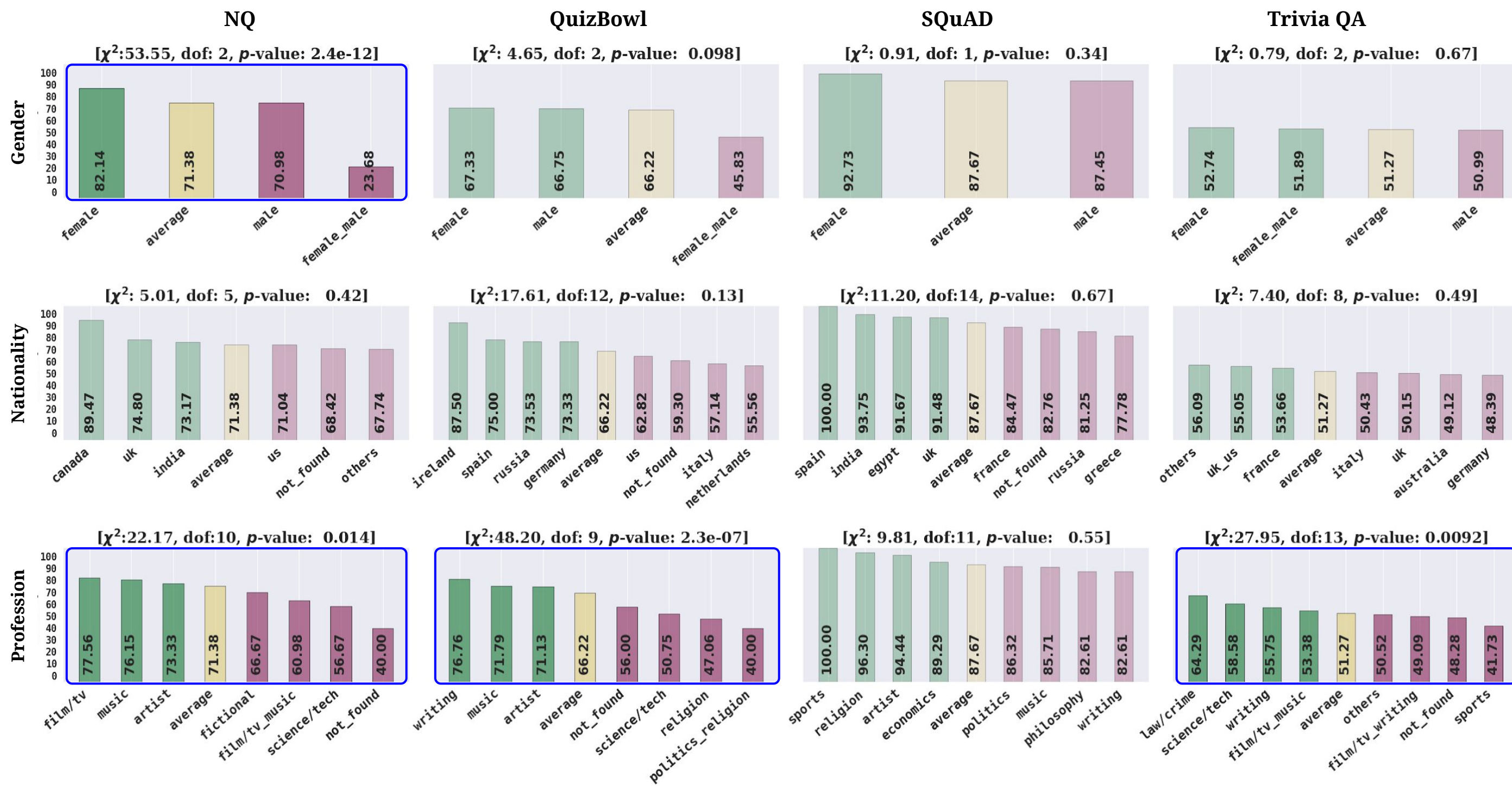**Accuracy** *(y-axis label for all subplots)*

Gender based Contingency Table for **NQ** examples

Gender values / Exact Match score	EM = 1 correct prediction	EM = 0 incorrect prediction
male	433	177
female	161	35
female_male	9	29

dof. = 2,
χ^2 = 53.55,
p-value = 2.4e-12