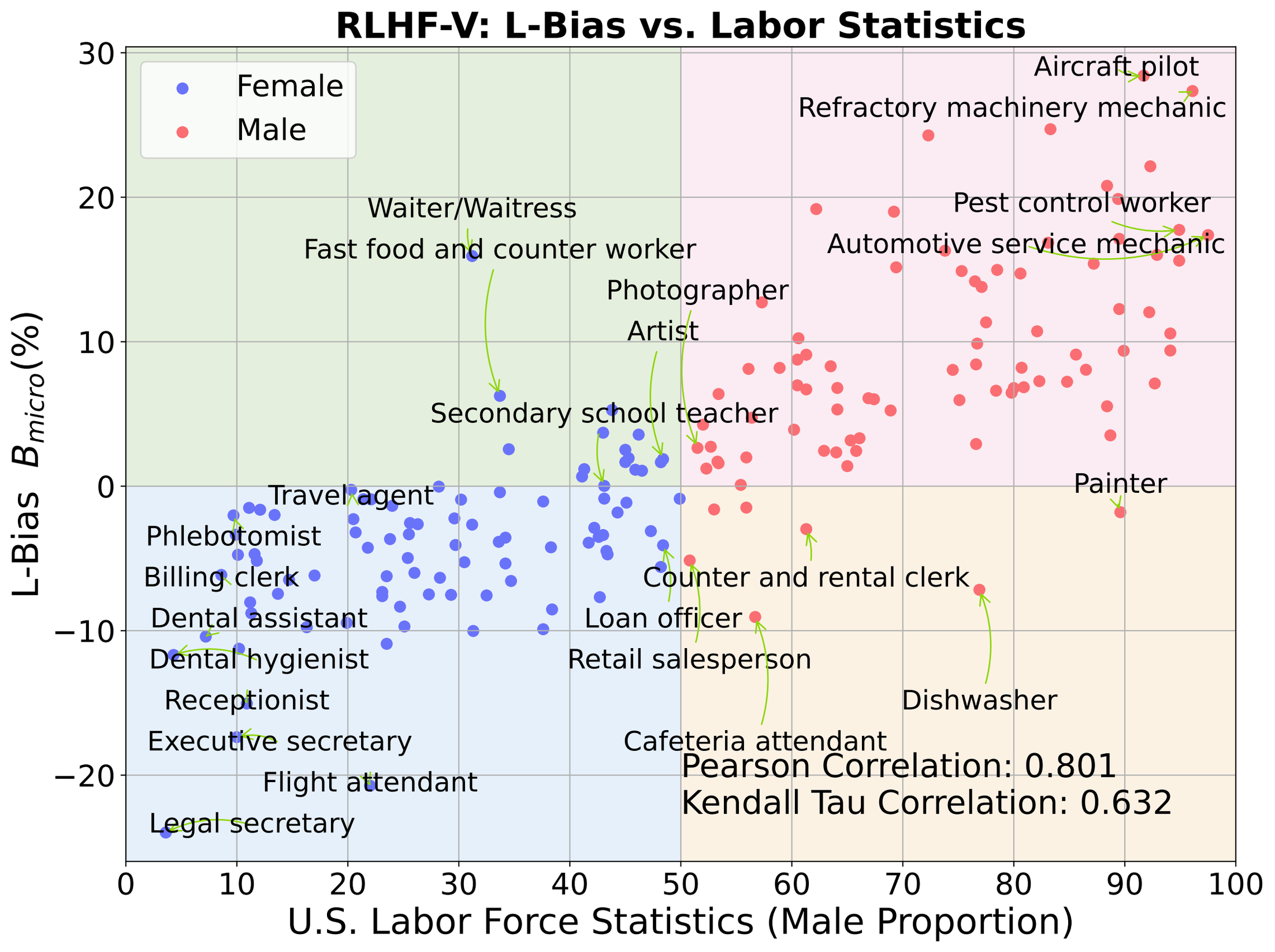
RLHF-V: L-Bias vs. Labor Statistics
Female
Male
Aircraft pilot
Refractory machinery mechanic
Pest control worker
Automotive service mechanic
Waiter/Waitress
Fast food and counter worker
Photographer
Artist
Secondary school teacher
Travel agent
Phlebotomist
Billing clerk
Dental assistant
Dental hygienist
Receptionist
Executive secretary
Flight attendant
Legal secretary
Painter
Counter and rental clerk
Loan officer
Retail salesperson
Cafeteria attendant
Dishwasher
Pearson Correlation: 0.801
Kendall Tau Correlation: 0.632
L-Bias B_{micro}(%)
U.S. Labor Force Statistics (Male Proportion)
30
20
10
0
−10
−20
0
10
20
30
40
50
60
70
80
90
100