Trends in Researchers' Diversity over a Ten-Year Period in Three Nature Language Process Conferences

Wenbo Shi

August 1, 2020

Introduction

For this analysis, I chose the Nature Language Process (NLP) domain because it is very important for humans and computers to communicate and can help computers be more understand about natural human language. NLP is one of the branches under the Artificial Intelligence (AI), so it is widely used in our life now, such as translation, text recognition and voice assistant. In this analysis, we will discover how researchers' diversity changes in three aspects through three conferences: population, gender and race. The three conferences are Association for Computational Linguistics (ACL), Empirical Methods in Natural Language Processing (EMNLP) and North American Chapter of the Association for Computational Linguistics (NAACL).

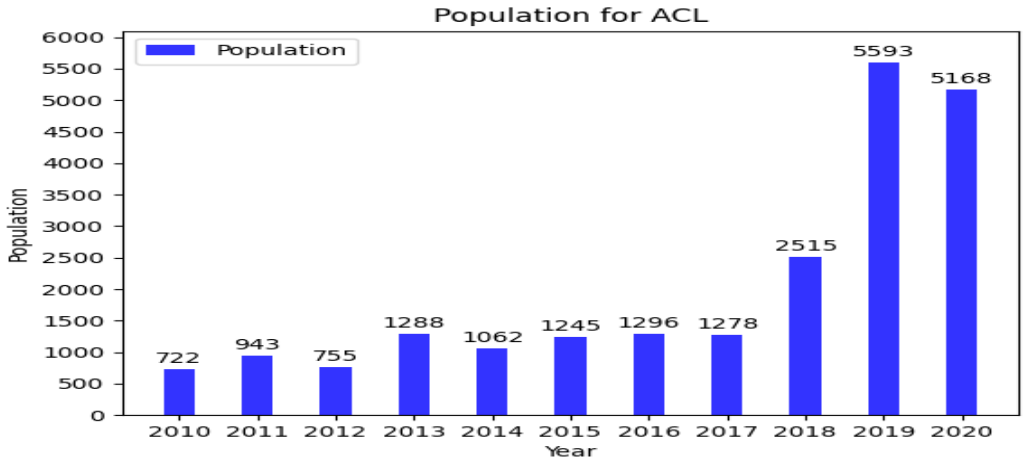
Methods

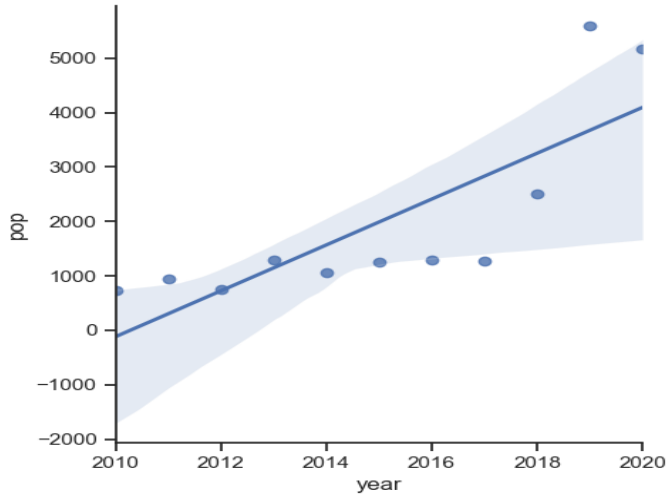
We will see the diversity change by analyzing the names of researchers in accept papers for three conferences. I used BeautifulSoup library to crawl the website and got all the names of researchers, which could help me draw the population bar chart and the linear regression from the first year to the last year of every conference. Then, using NamSor API to analyze the gender and race of the names so that I could draw the bar charts for the gender and race. In terms of the test method, I used hypothesis test to assume the

proportion change in gender and race of every conference between 2010 and 2019, which based on the significance level (Alpha) equals to 5%. The goal of the hypothesis test was for finding the change of researchers' diversity and I used two tailed test. The reason to set 2019 year instead of 2020 year is there was an external influence in 2020, which was the Covid-19. If using the data from 2020, it may have some inaccuracy and it does not have the reference value. Besides, EMNLP and NAACL did not have any data in 2020 since the Covid-19 blocked the progress of the conference, and NAACL conference did not show the data for 2011, 2014 and 2017 since NAACL lost all of the data for the three years. Therefore, the analysis would exclude the three years. After setting up my hypothesis test, I used z test to determine z value and p value for getting the results. If p value is bigger than Alpha, we fail to reject the null hypothesis (H0) and if p value is smaller than Alpha, we can reject the null hypothesis (H0). I chose to use z test because of the huge sample size, and it could get a more accurate result if using z test.

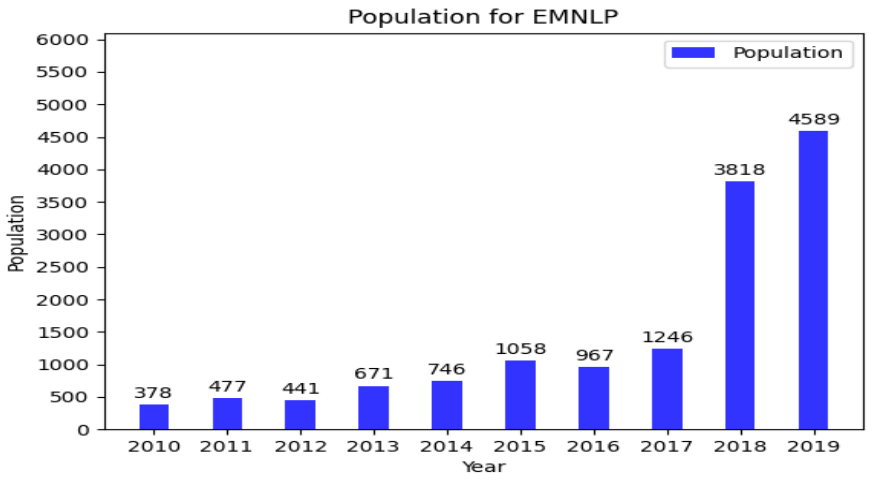
Analysis

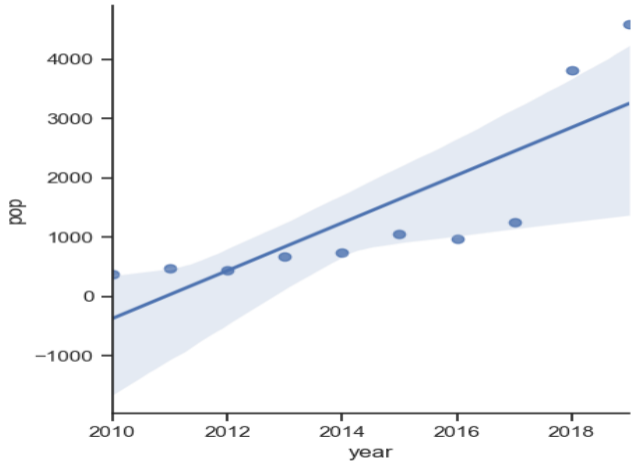
1. Population



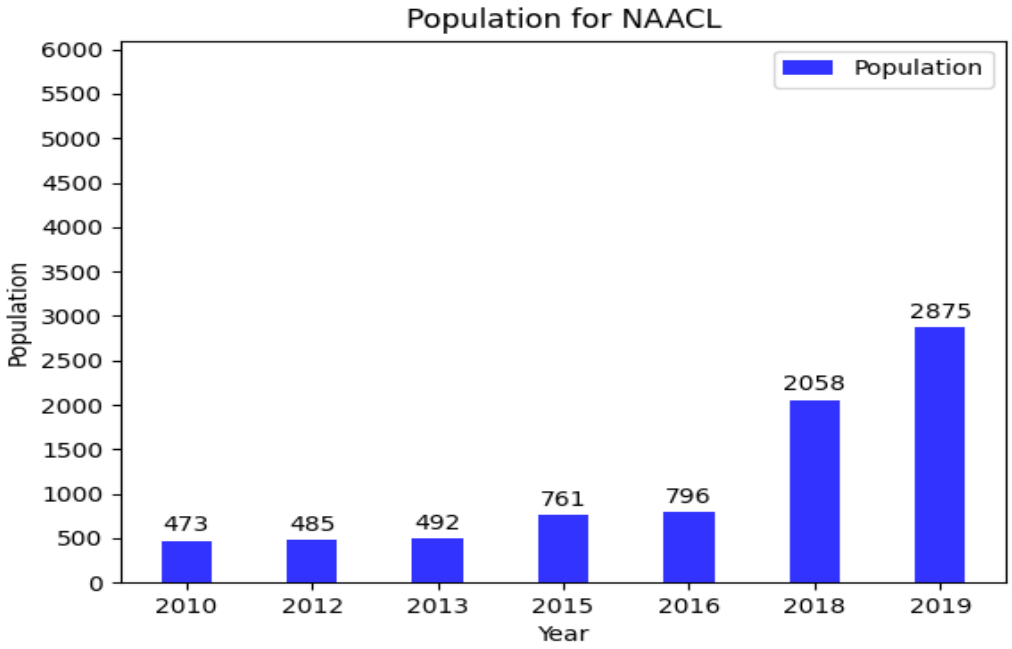


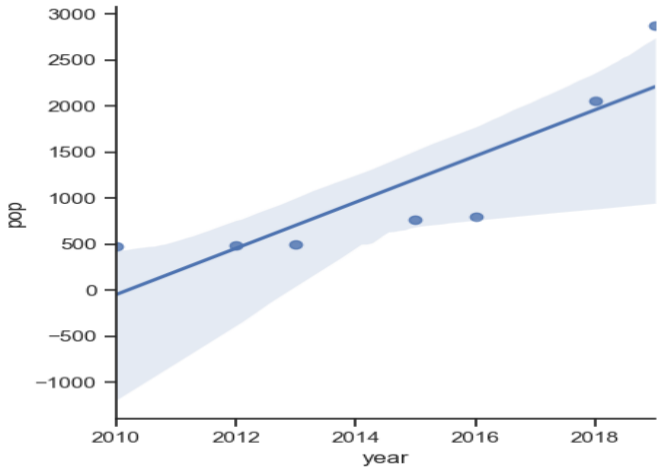
As we can see from the bar chart, in the ACL conference, the largest number of population was in 2019 and the smallest number of population was in 2010. One sharply rose was between 2018 and 2019, which varied from numbers of 2515 to 5593. Besides, from the linear regression of the population in ACL, we can see that there is a gradual increase from 2010 to 2020.





In this population bar chart for EMNLP, the smallest number of population was in 2010 which was 378 and the largest number of population was in 2019 which was 4589. From the linear regression of the population in EMNLP, we can see the general trend was gradually increasing from 2010 to 2019.





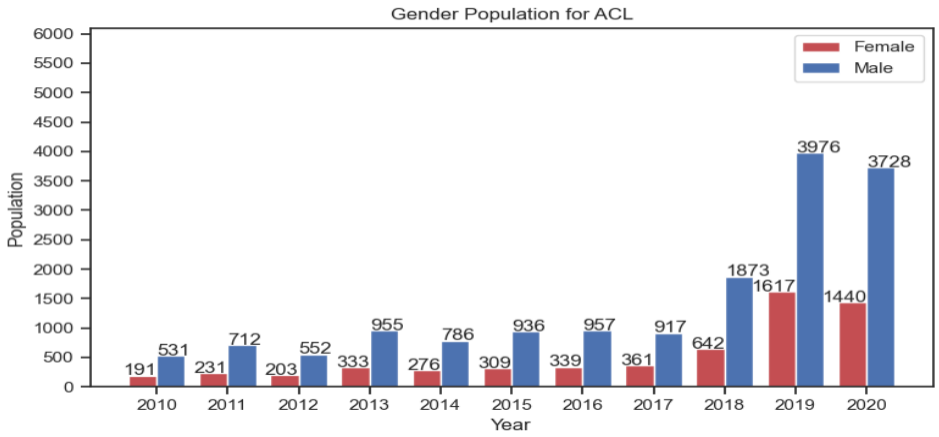
In terms of this population bar chart for NAACL, in 2010, the number of population took up 473 as the smallest and the largest number was in 2019, which occupied 2875. As we can see from the linear regression of the population in NAACL, the whole trend was gradually increasing from 2010 to 2019.

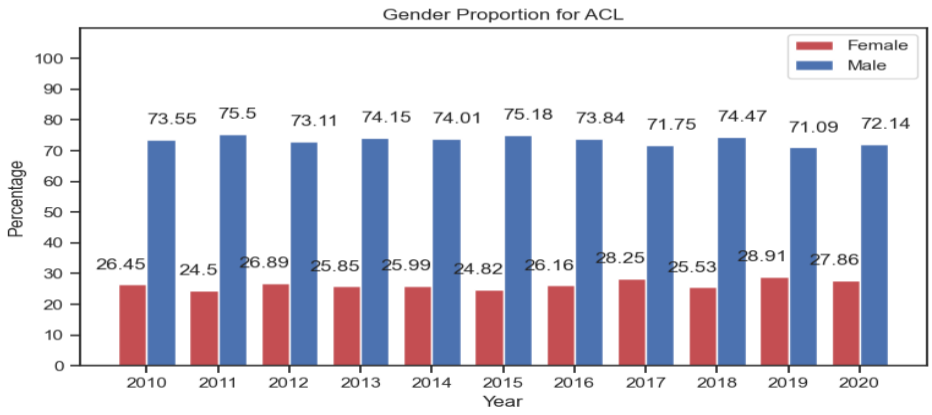
2. Gender

When we need to see the diversity change of gender in each conference, the first thing is to use hypothesis test. For all of the three conferences, my null hypothesis was female proportion did not change in 2010 and 2019 (H0: year1 == year2) and the alternative hypothesis was female proportion did change in 2010 and 2019 (HA: year1 != year2). The reason to test the change of female proportion for the 3 conferences is the female percentage was smaller than male percentage, through the change of female proportion, we could get the change of gender diversity.

For the ACL conference, my input values were the population (2010: 722, 2019: 5593)

and the number of females (2010: 191, 2019: 1617), then I got the z value: 1.3743979668351824 and the p value: 0.16931816701420466. Since the p value was greater than Alpha, we could get the conclusion that failed to reject H0, and female proportion did not change in 2010 and 2019, which suggests the gender diversity of ACL conference did not have statistically change.

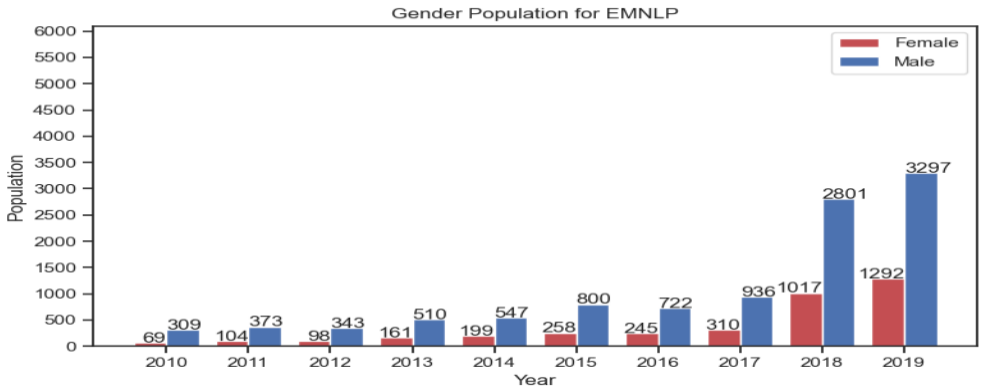


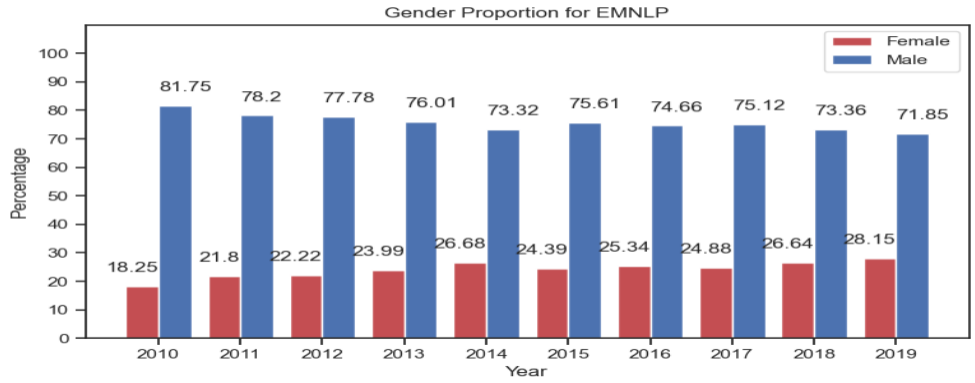


As we can see from the gender bar chart of ACL, the female proportion during 2011 and 2018 did not have a sharp change. (2011: 24.5%, 2018: 25.53%). Since the female proportion for ACL between 2011 and 2018 had always been average, and the change of male proportion was also at an average level (2011: 75.5%, 2018: 74.47%). In

conclusion, the diversity of gender almost did not change from time to time.

For the EMNLP conference, my input values were the population (2010: 378, 2019:4589) and the numbers of female (2010: 69, 2019: 1292), then I got the z value: 4.148193836411904 and the p value: 3.351086162621253e-05. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. Therefore, for the diversity of gender, EMNLP conference had a change.

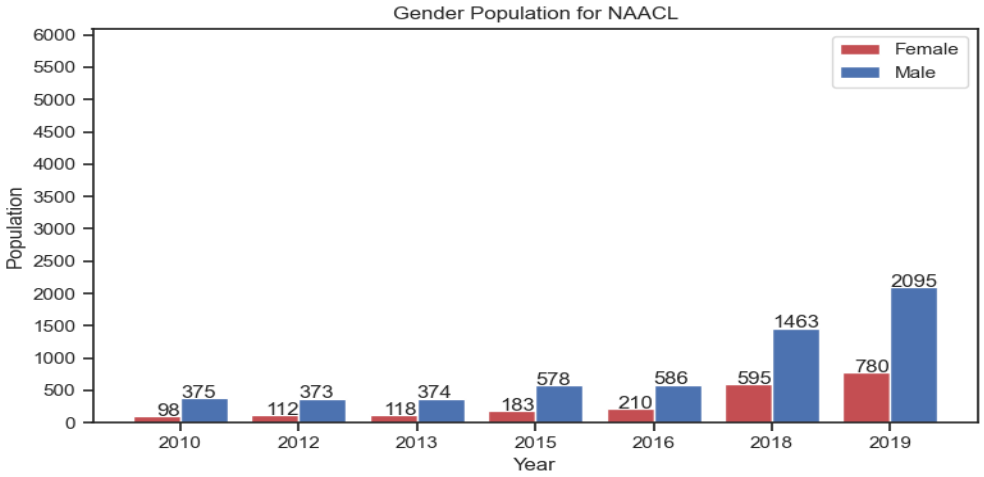


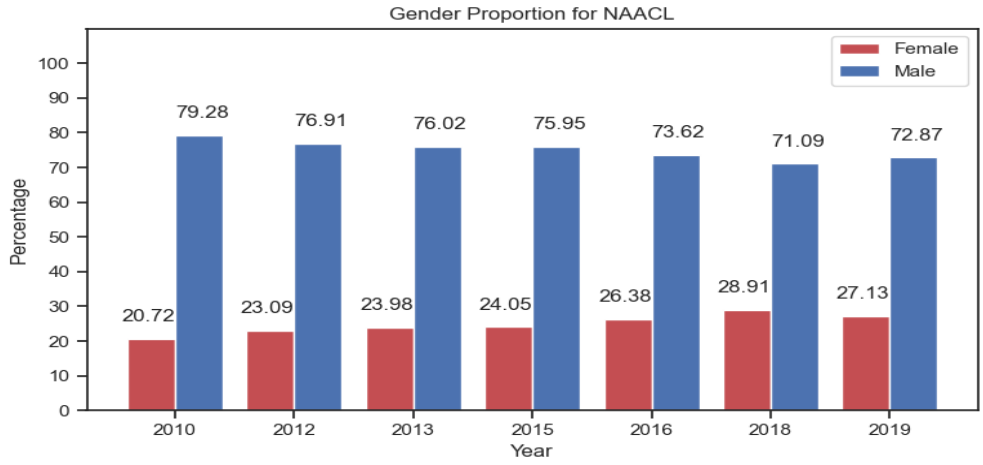


From this gender bar chart of EMNLP, we can visually see the female proportion during 2011 and 2018 did not have big fluctuations (2011: 21.8%, 2012: 22.22%, 2013: 23.99%, 2014: 26.68%, 2015: 24.39%, 2016: 25.34%, 2017: 24.88%, 2018: 26.64%) but there was

an huge increase through 2010 and 2019 (2010: 18.25%, 2019: 28.15%), which had a contribution to the gender diversity.

Regarding to the NAACL conference, my input values were the population (2010: 473, 2019: 2875) and the numbers of female (2010: 98, 2019: 780), then I got the z value: 2.937745521416374 and the p value: 0.0033060827543516303. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. Therefore, in terms of the diversity of gender, NAACL conference changed.





In this gender bar chart of NAACL, a slight increase of female proportion from 2012 to

2018 could be observed (2012: 23.09%, 2013: 23.98%, 2015: 24.05%, 2016: 26.38%, 2018: 28.91%), and it had a significant rise through 2010 and 2019 (2010: 20.72%, 2019: 27.13%). Therefore, the diversity of gender changed because of the change in female proportion.

By observing the data, in three conferences, I found that female proportion was generally lower than male proportion though it indeed had increased in 2010 and 2019. The female proportion of EMNLP changed from 18.25% to 28.15% in 2010 and 2019, which rose around 10%. NAACL also had a female proportion change from 20.72% to 27.13% in 2010 and 2019, which increased around 7%. For ACL, although we did not get a statistically change, the gender bar chart showed the very slight change of female proportion in 2010 and 2019, which varied from 26.45% to 28.91%. It indicated the female researchers gradually increased in the area of NLP for three conferences, so that the gender diversity was improved in this area.

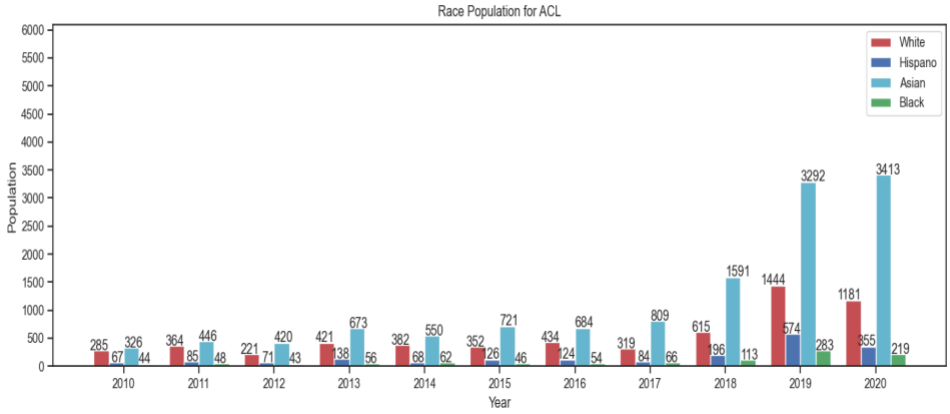
3. Race

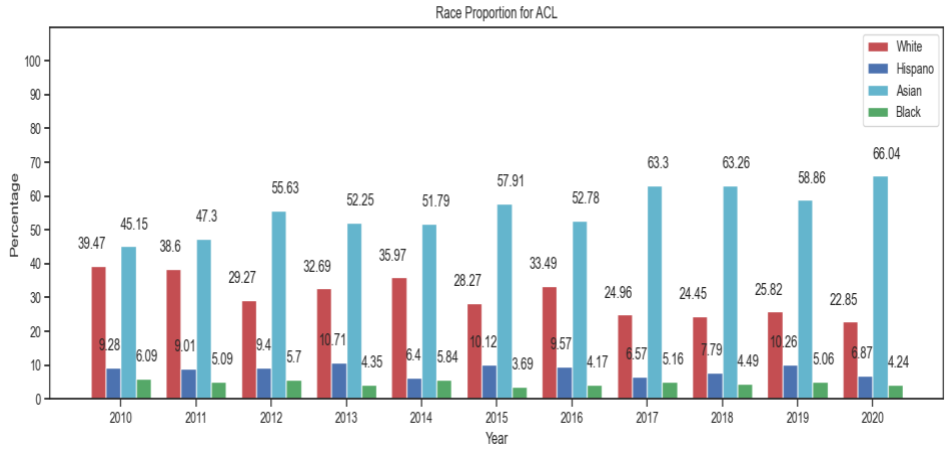
To test the change of race proportion for three conferences, I divided into 4 types of race, which were White, Hispano, Asian and Black. For all 3 conferences, I set up the same null hypothesis and alternative hypothesis for each type of race. For the White, the null hypothesis was the White proportion does not change in 2010 and 2019 (H0: year1 == year2) and the alternative hypothesis was the White proportion does change in 2010 and 2019 (HA: year1 != year2). For the Hispano, the null hypothesis was the Hispano proportion does not change in 2010 and 2019 (H0: year1 == year2) and the alternative

hypothesis was the Hispano proportion does change in 2010 and 2019 (HA: year1 != year2). For the Asian, the null hypothesis was the Asian proportion does not change in 2010 and 2019 (H0: year1 == year2) and the alternative hypothesis was the Asian proportion does change in 2010 and 2019 (HA: year1 != year2). For the Black, the null hypothesis was the Black proportion does not change in 2010 and 2019 (H0: year1 == year2) and the alternative hypothesis was the Black proportion does change in 2010 and 2019 (HA: year1 != year2).

To test the change of White proportion for the ACL conference, my input values were the population (2010: 722, 2019: 5593) and the numbers of the White (2010: 285, 2019: 1444), then I got the z value: -7.7442087288059005 and the p value: 9.617890118029388e-15. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. The next was to test the change of Hispano proportion in ACL, my input values were the population (2010: 722, 2019: 5593) and the numbers of the Hispano (2010: 67, 2019: 574), then I got the z value: 0.8231513140347424 and the p value: 0.4104219511068028. Since the p value was greater than Alpha, we could get the conclusion which failed to reject H0. For testing the change of Asian proportion in ACL, my input values were the population (2010: 722, 2019: 5593) and the numbers of the Asian (2010: 326, 2019: 3292), then I got the z value: 7.007179576153298 and the p value: 2.4317008301557492e-12. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. Regarding to the last part, it was to test the change in Black proportion in ACL, my input values were the population (2010: 722, 2019: 5593) and the numbers of the Black (2010: 44, 2019: 283), then I got the z value: -

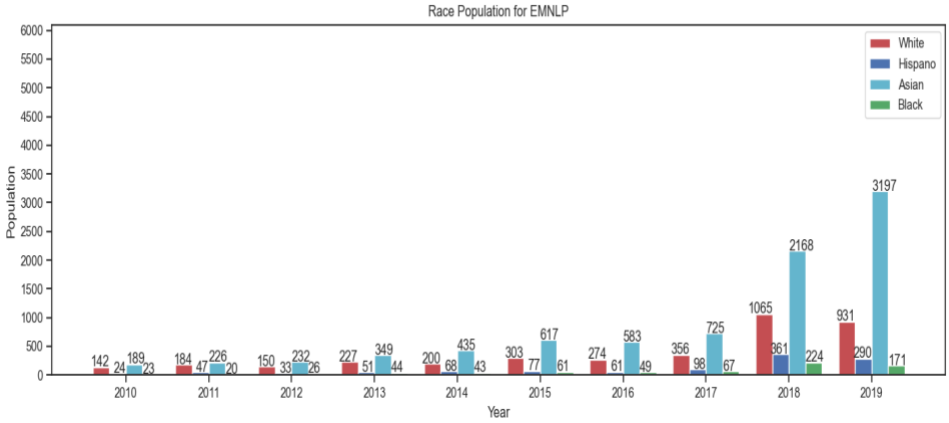
1.1803309115889864 and the p value: 0.23786862838348977. Since the p value was greater than Alpha, we could get the conclusion which failed to reject H0. Therefore, as the results indicated above, White proportion and Asian proportion of ACL changed in 2010 and 2019 so the race diversity had a change.

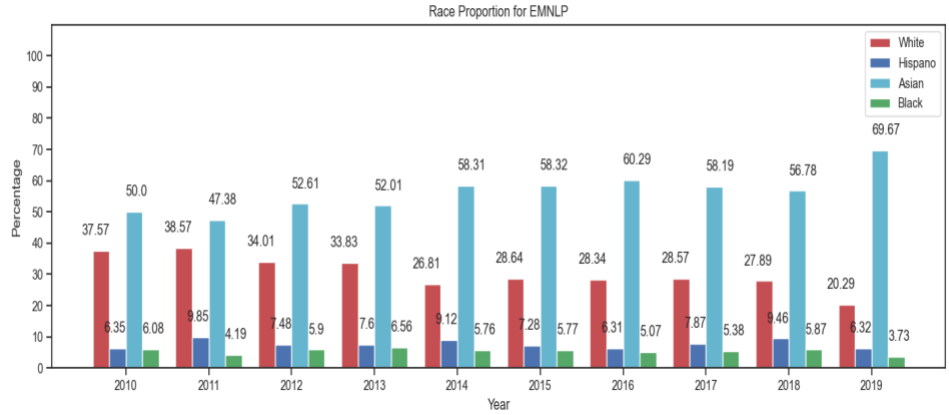




From this race bar chart for ACL, from 2011 to 2018, Black proportion was always the lowest and Asian proportion was always the highest in the four races. Besides, White proportion almost stayed the same through 2017, 2018 and 2019. All in all, in ACL, the proportion of Asian researchers almost took up more than half from 2010 to 2019 and the proportion of Black researchers occupied lowest.

For testing the White proportion change in the EMNLP conference, my input values were the population (2010: 378, 2019: 4589) and the numbers of the White (2010: 142, 2019: 931), then I got the z value: -7.846212040669402 and the p value: 4.287917118865662e-15. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. To test Hispano proportion change, my input values were the population (2010: 378, 2019: 4589) and the numbers of the Hispano (2010: 24, 2019: 290), then I got the z value: -0.022843387271168532 and the p value: 0.9817751990056582. Since the p value was greater than Alpha, we could get the conclusion which failed to reject H0. Testing the Asian proportion change of EMNLP, my input values were the population (2010: 378, 2019: 4589) and the numbers of the Asian (2010: 189, 2019: 3197), then I got the z value: 7.889904559697434 and the p value: 3.0241772339953938e-15. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. Lastly, I tested the Black proportion change, my input values were the population (2010: 378, 2019: 4589) and the numbers of the Black (2010: 23, 2019: 171), then I got the z value: -2.2749111148300973 and the p value: 0.022911262451648443. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. In summary, the results of testing statistically showed the proportion of White, Asian as well as Black changed in 2010 and 2019, and the Hispano proportion did not change in 2010 and 2019.

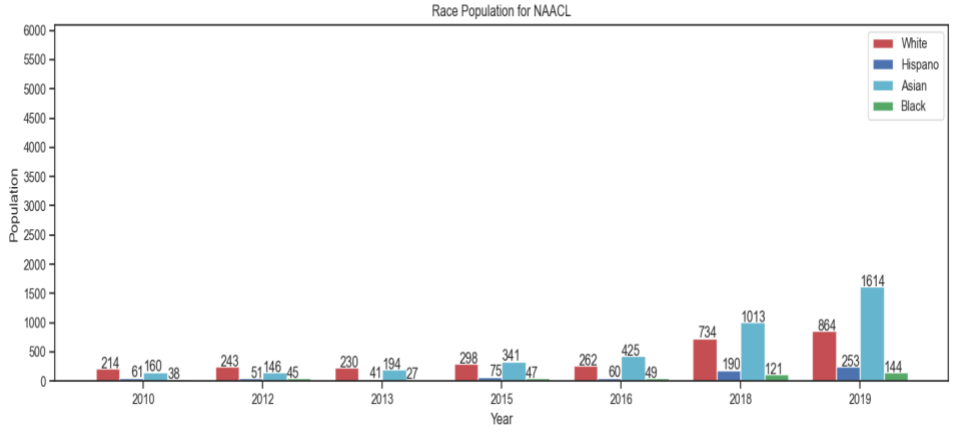


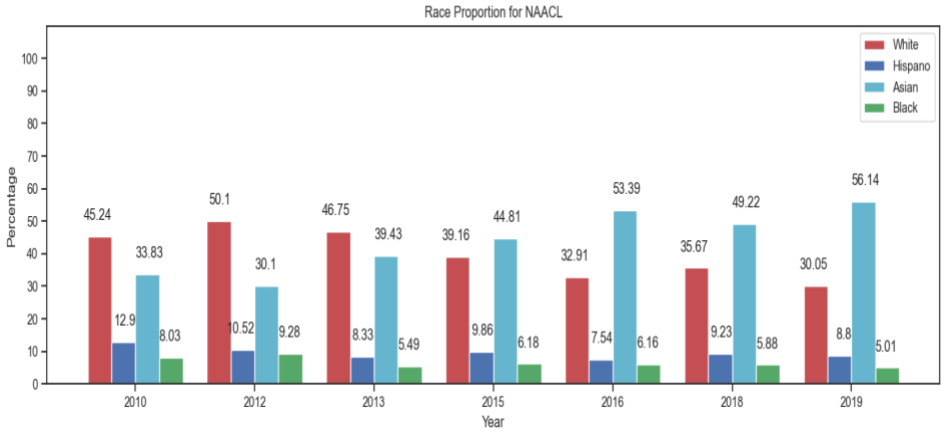


As we can see from the race bar chart for EMNLP, the Asian researchers occupied the largest proportion every year in EMNLP. The Black researchers took up the smallest proportion every year and the Black proportion did not have great change between 2010 and 2019. In 2010 and 2019, the White proportion fell down from 37.57% to 20.29% and Asian proportion rose from 50% to 69.67%, which decreased the level of race diversity, because there were more Asian researchers and less White researchers.

When talking about the NAACL conference, to test the White proportion change, my input values were the population (2010: 473, 2019: 2875) and the numbers of the White (2010: 214, 2019: 864), then I got the z value: -6.552469215049982 and the p value:

5.659338822969305e-11. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. To test the Hispano proportion change, my input values were the population (2010: 473, 2019: 2875) and the numbers of the Hispano (2010: 61, 2019: 253), then I got the z value: -2.8318653026501823 and the p value: 0.004627733869533111. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. For testing the Asian proportion change, my input values were the population (2010: 473, 2019: 2875) and the numbers of the Asian (2010: 160, 2019: 1614), then I got the z value: 9.009716479609654 and the p value: 2.0658988151606658e-19. Since the p value was less than Alpha, we could get the conclusion which was to reject H0. For the last one, to test the Black proportion change, my input values were the population (2010: 473, 2019: 2875) and the numbers of the Black (2010: 38, 2019: 144), then I got the z value: -2.6890275662513434 and the p value: 0.007166050126334939. Since the p value was smaller than Alpha, we could get the conclusion which was to reject H0. In summary, when testing the proportion change of all the race types in NAACL, the results demonstrated all of them had changed in 2010 and 2019.





By observing this race bar chart of NAACL, from 2010 to 2013, the proportion of White researchers was the highest, but the proportion of Asian researchers became dominant between 2015 and 2019. The proportion of Hispano and Black had a slight change in 2010 and 2019. The White proportion decreased from 45.24% to 30.05% and the Asian proportion increased from 33.83% to 56.14% through 2010 to 2019. Therefore, the race diversity was unbalanced.

Comparison

After doing the longitudinal analysis, I started the horizontal comparison among three conferences through the hypothesis test and z test.

1. ACL and EMNLP conference

When comparing the ACL and EMNLP conference, the proportion of four races were statistically same in 2010. However, from the gender bar chart of the two conferences, in 2010, we could see the female proportion of ACL occupied 26.45% and female proportion of EMNLP took up 18.25%, which had a huge difference. Since female proportion was always lower than male proportion, the ACL conference contributed more

to the balance of gender diversity. In 2019, female proportion did not change between ACL and EMNLP. However, four races proportion changed between ACL and EMNLP in 2019. The biggest difference is the difference in the Asian proportion. In ACL of 2019, Asian proportion made up 58.86%, and in EMNLP of 2019, Asian proportion took up 69.67%, which means ACL was more diverse in terms of race because the distribution of each race proportion was more balanced when comparing with EMNLP based on the bar charts in 2019.

2. ACL and NAACL conference

To compare the female proportion of 2010 between ACL and NAACL, statistical test showed there was a difference. From the gender bar charts, we could also visually see the change. In 2010, female proportion took up 26.45% in ACL and 20.72% in NAACL. Therefore, ACL contributes more to the gender diversity. For comparing the four races proportion change between ACL and NAACL in 2010, only Asian proportion showed the difference statistically. From the race bar charts, Asian proportion occupied 45.15% in ACL and 33.83% in NAACL. Besides, by observing the race data from ACL and NAACL in 2010, we could get the conclusion that NAACL was more balanced in the race diversity. Comparing the female proportion between ACL and NAACL in 2019, statistical test did not show the change. For comparing the four races proportion in 2019, White and Asian proportion had a change between ACL and NAACL. From the race bar charts, the White proportion took up 25.82% in ACL and 30.05% in NAACL. Besides, the Asian proportion occupied 58.86% in ACL and 56.14% in NAACL. Therefore, through the four races data from race bar charts of ACL and NAACL, ACL had more

balanced race diversity.

3. EMNLP and NAACL conference

When comparing the female proportion between EMNLP and NAACL in both 2010 and 2019, the statistical test showed no change. Comparing the four races proportion of 2010 for both conferences, the statistical test indicated the proportion of White, Hispano and Asian changed between EMNLP and NAACL but Black proportion did not change. The White proportion occupied 37.57% in EMNLP and 45.24% in NAACL. The Hispano proportion took up 6.35% in EMNLP and 12.9% in NAACL. The Asian proportion made up 50% in EMNLP and 33.83% in NAACL. Through these data from the race bar charts between the two conferences, we could get the conclusion NAACL had more balanced race diversity. In 2019, all of the four races proportion between EMNLP and NAACL changed statistically. The White proportion occupied 20.29% in EMNLP and 30.05% in NAACL. The Hispano proportion took up 6.32% in EMNLP and 8.8% in NAACL. The Asian proportion made up 69.67% in EMNLP and 56.14% in NAACL. The Black proportion occupied 3.73% in EMNLP and 5.01% in NAACL. From the race bar charts, we could get the conclusion that NAACL was more diverse in terms of race in 2019 when comparing with EMNLP conference.

Conclusion

Overall, among the three conferences, the ACL and NAACL conference have more balanced diversity than EMNLP. However, in the Nature Language Process field, through the longitudinal analysis, the race part for the three conferences are not becoming more

diverse because the Asian researchers are always the dominant except the NAACL conference from 2010 to 2013, which the proportion of White researchers was the highest. Besides, the Asian proportion had a significant increase in 2010 and 2019 for three conferences. In terms of gender, all the three conferences become more diverse. In summary, more and more female researchers participate in the Nature Language Process field and Asian researchers are most focused on the work of Nature Language Process. The Black researchers as well as the Hispano researchers do not very focus on this Nature Language Process area.

Formula:

```
def hyp_test(size1, year1, size2, year2):
        print("Null Hypothesis H0:    year1 == year2")
        print("Alternative Hypothesis HA: year1 != year2")
        alpha = 0.05/2
        p_h = (year2 + year1)/(size2 + size1)
        stdd = ((p_h * (1 - p_h))/size2) + ((p_h * (1 - p_h))/size1)
        z_value = (year2/size2 - year1/size1) / math.sqrt(stdd)
        print("z value: ", z_value)
        p_value = scipy.stats.norm.sf(abs(z_value))*2
        print("p value: ", p_value)
        if p_value > alpha:
                print("Fail to reject H0: year1 == year2")
        else:
                print("Reject H0: year1 != year2")
```

Test result:

Female proportion did not change in 2010 and 2019 in ACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 1.3743979668351824

p value: 0.16931816701420466

Fail to reject H0: year1 == year2

White participants proportion did not change in 2010 and 2019 in ACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -7.7442087288059005

p value: 9.617890118029388e-15

Reject H0: year1 != year2

Hispano participants proportion did not change in 2010 and 2019 in ACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 0.8231513140347424

p value: 0.4104219511068028

Fail to reject H0: year1 == year2

Asian participants proportion did not change in 2010 and 2019 in ACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 7.007179576153298

p value: 2.4317008301557492e-12

Reject H0: year1 != year2

Black participants proportion did not change in 2010 and 2019 in ACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -1.1803309115889864

p value: 0.23786862838348977

Fail to reject H0: year1 == year2

****

Female proportion did not change in 2010 and 2019 in EMNLP

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 4.148193836411904

p value: 3.351086162621253e-05

Reject H0: year1 != year2

White participants proportion did not change in 2010 and 2019 in EMNLP

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -7.846212040669402

p value: 4.287917118865662e-15

Reject H0: year1 != year2

Hispano participants proportion did not change in 2010 and 2019 in EMNLP

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -0.022843387271168532

p value: 0.9817751990056582

Fail to reject H0: year1 == year2

Asian participants proportion did not change in 2010 and 2019 in EMNLP

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 7.889904559697434

p value: 3.0241772339953938e-15

Reject H0: year1 != year2

Black participants proportion did not change in 2010 and 2019 in EMNLP

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -2.2749111148300973

p value: 0.022911262451648443

Reject H0: year1 != year2

****

Female proportion did not change in 2010 and 2019 in NAACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 2.937745521416374

p value: 0.0033060827543516303

Reject H0: year1 != year2

White participants proportion did not change in 2010 and 2019 in NAACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -6.552469215049982

p value: 5.659338822969305e-11

Reject H0: year1 != year2

Hispano participants proportion did not change in 2010 and 2019 in NAACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -2.8318653026501823

p value: 0.004627733869533111

Reject H0: year1 != year2

Asian participants proportion did not change in 2010 and 2019 in NAACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 9.009716479609654

p value: 2.0658988151606658e-19

Reject H0: year1 != year2

Black participants proportion did not change in 2010 and 2019 in NAACL

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -2.6890275662513434

p value: 0.007166050126334939

Reject H0: year1 != year2

****

Female proportion does not change between ACL and EMNLP in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -3.040294186394996

p value: 0.0023634716535820225

Reject H0: year1 != year2

White participants proportion does not change between ACL and EMNLP in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -0.616544656335265

p value: 0.5375351083322292

Fail to reject H0: year1 == year2

Hispano participants proportion does not change between ACL and EMNLP in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -1.6757026805926514

p value: 0.09379644451642045

Fail to reject H0: year1 == year2

Asian participants proportion does not change between ACL and EMNLP in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 1.5302420802868866

p value: 0.12595681896085226

Fail to reject H0: year1 == year2

Black participants proportion does not change between ACL and EMNLP in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -0.006274333943270438

p value: 0.9949938386638459

Fail to reject H0: year1 == year2

****

Female proportion does not change between ACL and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -2.2643298229998856

p value: 0.023553836638294065

Reject H0: year1 != year2

White participants proportion does not change between ACL and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 1.9777083548065164

p value: 0.04796161720160821

Fail to reject H0: year1 == year2

Hispano participants proportion does not change between ACL and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 1.976968323038355

p value: 0.04804521091977256

Fail to reject H0: year1 == year2

Asian participants proportion does not change between ACL and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -3.8976922031175265

p value: 9.711373887355108e-05

Reject H0: year1 != year2

Black participants proportion does not change between ACL and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 1.2970317249448498

p value: 0.19462027126495318

Fail to reject H0: year1 == year2

****

Female proportion does not change between EMNLP and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 0.8995909419284366

p value: 0.36833797940203694

Fail to reject H0: year1 == year2

White participants proportion does not change between EMNLP and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 2.2558213665097306

p value: 0.02408182065962256

Reject H0: year1 != year2

Hispano participants proportion does not change between EMNLP and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 3.1649946618221327

p value: 0.0015508579639291439

Reject H0: year1 != year2

Asian participants proportion does not change between EMNLP and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: -4.766251694622806

p value: 1.8768477775186738e-06

Reject H0: year1 != year2

Black participants proportion does not change between EMNLP and NAACL in 2010

Null Hypothesis H0: year1 == year2

Alternative Hypothesis HA: year1 != year2

z value: 1.0952484041085444

p value: 0.2734078249816301

Fail to reject H0: year1 == year2

****