1. **<u>What does the landscape of advisor and agent careers look like?</u>**

- How do people move between channels (BD, RIA, Dually registered)—counts, percentages?

Ans:

I merge data from BD_2013_12 to BD_2020_12, and RIA_2013_12 to RIA_2020_12 to solve these questions by using the unique identifier RepCRD. Considering eight periods as a whole, I find there are 45471 individuals moving from BD to Dually registered. The percentage is 25.18%. Additionally, there are 2552 individuals moving from RIA to Dually registered. The percentage is 5.15%.

2. **What are the rates of leave (join/stay) in RIA channel ?**

Period	number_stay	rate_stay_RIA	rate_leave_RIA	rate_join_RIA
2013_2014	310409	0.8901816	0.10981838	0.14407002
2014_2015	317429	0.8752871	0.12471288	0.12801657
2015_2016	338865	0.9308685	0.06913148	0.09664426
2016_2017	350641	0.9347485	0.06525147	0.09391110
2017_2018	362023	0.9355010	0.06449896	0.09011099
2018_2019	372694	0.9367089	0.06329108	0.08384198
2019_2020	383003	0.9414997	0.05850035	0.07413621

As we can see, the rate of leaving RIA is increasing from 0.1098 to 0.1247 from 2013 to 2015. The rate of leaving RIA is decreasing from 0.1247 to 0.0585 from 2015 to 2020. The rate of joining RIA is decreasing from 0.1441 to 0.0741 from 2013 to 2020. The rate of staying in RIA is decreasing from 0.8902 to 0.8743 from 2013 to 2015 and increasing from 0.8753 to 0.9415 from 2015 to 2020.

3. **For people currently at RIAs, did they come from BDs (or large companies)?**

Ans:

Date	count_fromBD	percent_fromBD
2013_12	15185	0.04354709
2014_12	19911	0.05490312
2015_12	24168	0.06638995
2016_12	29328	0.07818340
2017_12	32031	0.08277108
2018_12	35306	0.08873619
2019_12	38250	0.09402632
2020_12	40733	0.09846714

I will consider this question periodically. The number of people currently at RIA coming from BD is 15185 in 2013, 19911 in 2014, 24168 in 2015, 29328 in 2016, 32031 in 2017, 35306 in 2018, 38250 in 2019, and 40733 in 2020. The percentage of people currently at RIA coming from BD is 0.0435 in 2013, 0.0549 in 2014, 0.0664 in 2015, 0.0782 in 2016, 0.0828 in 2017, 0.0887 in 2018, 0.0940 in 2019, 0.0985 in 2020. We can see the percentage of people currently at RIA coming from BD is increasing from 2013 to 2020.

4. **Once people become dually registered, do they typically stay? Become just RIAs, just BDs?**

Ans:

For BD, there are 21698 individuals moving from dually registered to BD. The percentage of people staying dually registered is 90.48%. The percentage of people becoming BD is 9.52%.

For RIA, there are 9844 individuals moving from dually registered to RIA. The percentage of people staying dually registered is 95.68%. The percentage of people becoming RIA is 4.32%.

5. **How many BDs/RIAs/Dually Registered also sell insurance?**

Ans:

There are 58743 BDs selling insurance, 9503 RIAs selling insurance and 166017 Dually Registered selling insurance.

6. **Do minorities leave BDs (or large companies) for RIAs at higher rates than overall?**

Ans:

I define the definition of minorities as race and gender.

Date	Male	Female	AfricanAmerican	AmericanIndian	Asian	Hispanic	Hawaiian	Other	White
2013_12	0.04692989	0.02377825	0.02577689	0.02390438	0.03436478	0.02927928	0.010989011	0.05197935	0.04426472
2014_12	0.05714509	0.03447732	0.03424104	0.04119850	0.04466324	0.03770762	0.009708738	0.06250000	0.05437970
2015_12	0.07167297	0.04442381	0.04079284	0.04620462	0.05518305	0.04662917	0.026315789	0.07310259	0.06574563
2016_12	0.08194695	0.05020446	0.04652539	0.04923077	0.06792568	0.05365127	0.023809524	0.08554913	0.07707501
2017_12	0.08698666	0.05429350	0.05034627	0.05507246	0.07021135	0.05824331	0.015873016	0.08993779	0.08285421
2018_12	0.09303068	0.05962860	0.05752440	0.06068602	0.07472407	0.06462001	0.044444444	0.09838123	0.08978976
2019_12	0.09900778	0.06350681	0.06332914	0.07980050	0.07777778	0.06902638	0.060000000	0.10345630	0.09586210
2020_12	0.10442431	0.06689434	0.06747798	0.10256410	0.08046998	0.07392097	0.054878049	0.10580980	0.10134466

From the data, the rate of Male leaving BDs for RIAs is generally higher than that of Female leaving BDs for RIAs. Male > Female.

The general ranking of the rate of race minorities leaving BDs for RIAs is:

Other > White > Asian/American Indian(before 2019, Asian is higher than American Indian; after 2019, American Indian is higher than Asian) > African American > Hispanic > Hawaiian.

7. **What proportion of individuals at RIAs from BDs were minorities (race / gender)? Are there disproportionately more from these areas?**

Ans:

Date	Male	Female	AfricanAmerican	AmericanIndian	Asian	Hispanic	Hawaiian	Other	White
2013_12	0.8005927	0.1156404	0.01679261	0.0005597537	0.04216811	0.03395839	9.329228e-05	0.02817427	0.8782536
2014_12	0.8445081	0.1428356	0.01787222	0.0007739938	0.04404728	0.03546299	7.036307e-05	0.02751196	0.8742612
2015_12	0.8442155	0.1445300	0.01765357	0.0007747648	0.04554510	0.03624792	1.660210e-04	0.02606530	0.8735473
2016_12	0.8407665	0.1448445	0.01731812	0.0007050632	0.04913409	0.03648702	1.321994e-04	0.02608734	0.8701362
2017_12	0.8385002	0.1475133	0.01756606	0.0007287232	0.04867104	0.03804702	7.670771e-05	0.02550531	0.8694051
2018_12	0.8341642	0.1521554	0.01849283	0.0007595271	0.04963345	0.03975959	1.981375e-04	0.02588997	0.8652665
2019_12	0.8345359	0.1536732	0.01895927	0.0009276707	0.05012321	0.04081751	2.609074e-04	0.02577185	0.8631396
2020_12	0.8341639	0.1553286	0.01909977	0.0011310763	0.05123262	0.04274954	2.313565e-04	0.02514074	0.8604149

From the data, the proportion of individuals at RIAs from BDs are male is generally higher than the proportion of individuals at RIAs from BDs are female.

The proportion of individuals at RIAs from BDs are White is greater than Asian greater than Hispanic greater than Other greater than African American greater than American Indian greater than Hawaiian.
(White > Asian > Hispanic > Other > African American > American Indian > Hawaiian)

8. What is the distribution of gender, and race over time @ top firms (BD, RIA, and Insurance)?

Ans:

I will consider UBS Financial Service Inc., Merrill Lynch, Pierce, Fenner & Smith Incorporated, and Edward Jones as top firms.

For UBS(BD):

Date	Male	Female	Total	M_Percent	F_Percent
2013_12	8081	3828	11999	0.6734728	0.3190266
2014_12	8081	3829	11957	0.6758384	0.3202308
2015_12	7778	3693	12000	0.6481667	0.3077500
2016_12	8031	3901	12050	0.6664730	0.3237344
2017_12	7989	4032	12219	0.6538178	0.3299779
2018_12	8030	4074	12370	0.6491512	0.3293452
2019_12	7701	3958	11913	0.6464367	0.3322421
2020_12	7415	3865	11530	0.6431049	0.3352125

(gender, male and female counts, and proportions)

Date	AmericanIndian	Asian	AfricanAmerican	Hispanic	White	Hawaiian	Other
2013_12	0.0009249743	0.04799589	0.02384378	0.07841727	0.8260021	1.027749e-04	0.02271326
2014_12	0.0007990412	0.04844187	0.02327207	0.07960447	0.8251099	9.988014e-05	0.02267279
2015_12	0.0008702379	0.04921679	0.02320634	0.08189905	0.8212145	1.933862e-04	0.02339973
2016_12	0.0008469791	0.05175983	0.02333898	0.08460380	0.8161114	3.764352e-04	0.02296254
2017_12	0.0010895224	0.05374977	0.02278918	0.08707100	0.8124206	3.631741e-04	0.02251680
2018_12	0.0010527239	0.05658391	0.02342311	0.08755154	0.8091938	2.631810e-04	0.02193175
2019_12	0.0011469914	0.05814364	0.02311629	0.08911241	0.8059820	2.646903e-04	0.02223399
2020_12	0.0011514615	0.06129318	0.02364925	0.08821966	0.8025686	2.657219e-04	0.02285208

(race, the proportion of the different races from different years)

For Merrill Lynch, Pierce, Fenner & Smith Incorporated(BD),

Date	Male	Female	Total	M_Percent	F_Percent
2013_12	20531	9077	30017	0.6839791	0.3023953
2014_12	20704	9253	30207	0.6854040	0.3063197
2015_12	19799	8904	31128	0.6360511	0.2860447
2016_12	21488	9510	31784	0.6760634	0.2992071
2017_12	22540	10067	33873	0.6654267	0.2971984
2018_12	22499	10176	34131	0.6591955	0.2981454
2019_12	19206	9272	29554	0.6498613	0.3137308
2020_12	19377	9374	29914	0.6477569	0.3133650

(gender, male and female counts, and proportions)

Date	AmericanIndian	Asian	AfricanAmerican	Hispanic	White	Hawaiian	Other
2013_12	0.0009458607	0.07107468	0.03247455	0.06891271	0.8041167	0.0003603279	0.02211512
2014_12	0.0009541157	0.07476798	0.03261341	0.06899991	0.7998525	0.0005204267	0.02229161
2015_12	0.0008933285	0.07552686	0.03382466	0.07166931	0.7956308	0.0005278759	0.02192715
2016_12	0.0008469356	0.07749461	0.03522482	0.07272097	0.7909224	0.0004619649	0.02232830
2017_12	0.0008772238	0.08182743	0.03386084	0.07684480	0.7830099	0.0004912453	0.02308853
2018_12	0.0009361731	0.08532549	0.03420375	0.08114614	0.7745160	0.0004680865	0.02340433
2019_12	0.0009450422	0.07811137	0.03591160	0.08814336	0.7736987	0.0004361733	0.02275371
2020_12	0.0008994672	0.08226666	0.03625545	0.09399433	0.7635093	0.0005535183	0.02252128

(race, the proportion of the different races from different years)

For Edward Jones(BD):

Date	Male	Female	Total	M_Percent	F_Percent
2013_12	11412	3547	15074	0.7570651	0.2353058
2014_12	12097	3803	15936	0.7590989	0.2386421
2015_12	11626	3653	16609	0.6999819	0.2199410
2016_12	12859	3977	17102	0.7519004	0.2325459
2017_12	14168	4484	19157	0.7395730	0.2340659
2018_12	15172	4927	20809	0.7291076	0.2367726
2019_12	15949	5225	21894	0.7284644	0.2386499
2020_12	15847	5278	21807	0.7266933	0.2420324

(gender, male and female counts, and proportions)

Date	AmericanIndian	Asian	AfricanAmerican	Hispanic	White	Hawaiian	Other
2013_12	0.001901612	0.01868541	0.03439438	0.03174866	0.8958247	0.0004960728	0.01694915
2014_12	0.001826762	0.01925712	0.03493682	0.03265337	0.8943523	0.0004566905	0.01651697
2015_12	0.001911775	0.01947178	0.03582808	0.03299582	0.8926574	0.0003540324	0.01678114
2016_12	0.001918624	0.01964935	0.03605690	0.03493219	0.8905723	0.0005292756	0.01634138
2017_12	0.002099615	0.02082118	0.03569346	0.03680159	0.8874956	0.0004665811	0.01662195
2018_12	0.002051876	0.02214973	0.03493450	0.03735466	0.8863577	0.0004208976	0.01673068
2019_12	0.001989905	0.02256843	0.03402252	0.03775966	0.8860415	0.0003882741	0.01722966
2020_12	0.001879523	0.02283620	0.03293863	0.03867118	0.8853961	0.0003759045	0.01790245

(race, the proportion of the different races from different years)

For UBS(RIA):

Date	Male	Female	Total	M_Percent	F_Percent
2013_12	7572	3138	10743	0.7048311	0.2920972
2014_12	7613	3190	10841	0.7022415	0.2942533
2015_12	6888	2796	10114	0.6810362	0.2764485
2016_12	7163	3031	10299	0.6955044	0.2943004
2017_12	6987	3002	10117	0.6906197	0.2967283
2018_12	6988	3054	10200	0.6850980	0.2994118
2019_12	6794	3004	9961	0.6820600	0.3015761
2020_12	6552	2973	9683	0.6766498	0.3070329

(gender, male and female counts, and proportions)

Date	AmericanIndian	Asian	AfricanAmerican	Hispanic	White	Hawaiian	Other
2013_12	0.0009024253	0.04839255	0.02346306	0.07422448	0.8309081	0.0001128032	0.02199662
2014_12	0.0007626934	0.04826760	0.02298976	0.07474395	0.8304642	0.0001089562	0.02266289
2015_12	0.0009030365	0.04876397	0.02257591	0.07438763	0.8303420	0.0002257591	0.02280167
2016_12	0.0008711750	0.05194381	0.02319503	0.07622781	0.8250027	0.0004355875	0.02232386
2017_12	0.0009770926	0.05222017	0.02290739	0.07751601	0.8237976	0.0004342634	0.02214743
2018_12	0.0011599705	0.05135506	0.02351577	0.08161974	0.8207318	0.0003163556	0.02130128
2019_12	0.0011550982	0.05271448	0.02289195	0.08306206	0.8183346	0.0003150268	0.02152683
2020_12	0.0012639562	0.05550874	0.02338319	0.08268380	0.8149357	0.0003159890	0.02190857

(race, the proportion of the different races from different years)

For Merrill Lynch, Pierce, Fenner & Smith Incorporated(RIA),

Date	Male	Female	Total	M_Percent	F_Percent
2013_12	17969	8107	26201	0.6858135	0.3094157
2014_12	17097	7748	24915	0.6862131	0.3109773
2015_12	16556	7598	25641	0.6456846	0.2963223
2016_12	17673	8104	26181	0.6750315	0.3095375
2017_12	18075	8143	26900	0.6719331	0.3027138
2018_12	18299	8286	27433	0.6670433	0.3020450
2019_12	17957	8308	27142	0.6615946	0.3060939
2020_12	18128	8431	27518	0.6587688	0.3063813

(gender, male and female counts, and proportions)

Date	AmericanIndian	Asian	AfricanAmerican	Hispanic	White	Hawaiian	Other
2013_12	0.0009968599	0.06335045	0.03309575	0.06968051	0.8106465	0.0003987440	0.02183123
2014_12	0.0010120433	0.06472017	0.03309382	0.06993219	0.8090274	0.0005566238	0.02165773
2015_12	0.0010460251	0.06627995	0.03432864	0.07255610	0.8040605	0.0005705591	0.02115824
2016_12	0.0009036281	0.06677811	0.03524149	0.07337460	0.8016536	0.0005421768	0.02150635
2017_12	0.0009432749	0.06915920	0.03485829	0.07803456	0.7941517	0.0005573897	0.02229559
2018_12	0.0010180397	0.07240298	0.03497984	0.08217616	0.7861302	0.0005293806	0.02276337
2019_12	0.0010236623	0.07720776	0.03574944	0.08705067	0.7758179	0.0004724595	0.02267806
2020_12	0.0009360141	0.08184507	0.03579318	0.09247819	0.7655472	0.0005990490	0.02280130

(race, the proportion of the different races from different years)

For Edward Jones(RIA),

Date	Male	Female	Total	M_Percent	F_Percent
2013_12	10261	2916	13253	0.7742398	0.2200257
2014_12	11046	3188	14263	0.7744514	0.2235154
2015_12	10761	3042	15060	0.7145418	0.2019920
2016_12	11990	3404	15633	0.7669673	0.2177445
2017_12	13165	3772	17361	0.7583089	0.2172686
2018_12	14196	4199	19021	0.7463330	0.2207560
2019_12	15317	4673	20634	0.7423185	0.2264709
2020_12	15372	4784	20777	0.7398566	0.2302546

(gender, male and female counts, and proportions)

Date	AmericanIndian	Asian	AfricanAmerican	Hispanic	White	Hawaiian	Other
2013_12	0.001776033	0.01897551	0.03430548	0.03327725	0.8943728	0.0005608525	0.01673210
2014_12	0.001695346	0.01983555	0.03450030	0.03348309	0.8939561	0.0005086039	0.01602102
2015_12	0.001853139	0.01945796	0.03520964	0.03389700	0.8926724	0.0003860706	0.01652382
2016_12	0.001937567	0.01951920	0.03573735	0.03473269	0.8912092	0.0005740940	0.01628992
2017_12	0.001914364	0.02016464	0.03484143	0.03726629	0.8888393	0.0005104971	0.01646353
2018_12	0.002057966	0.02166581	0.03441377	0.03818670	0.8865260	0.0004573258	0.01669239
2019_12	0.002052756	0.02181053	0.03366520	0.03828390	0.8865852	0.0004105512	0.01719183
2020_12	0.001921656	0.02202513	0.03242178	0.03936930	0.8858832	0.0003941858	0.01798473

(race, the proportion of the different races from different years)

9. How many title changes do BDs/RIAs/Insurance Agents have (over the course of the dataset)?

Ans:

For BD,

Var1	Freq
0	91434
1	156540
2	60722
3	25695
4	7905
5	1557
6	235
7	29
8	3

For RIA,

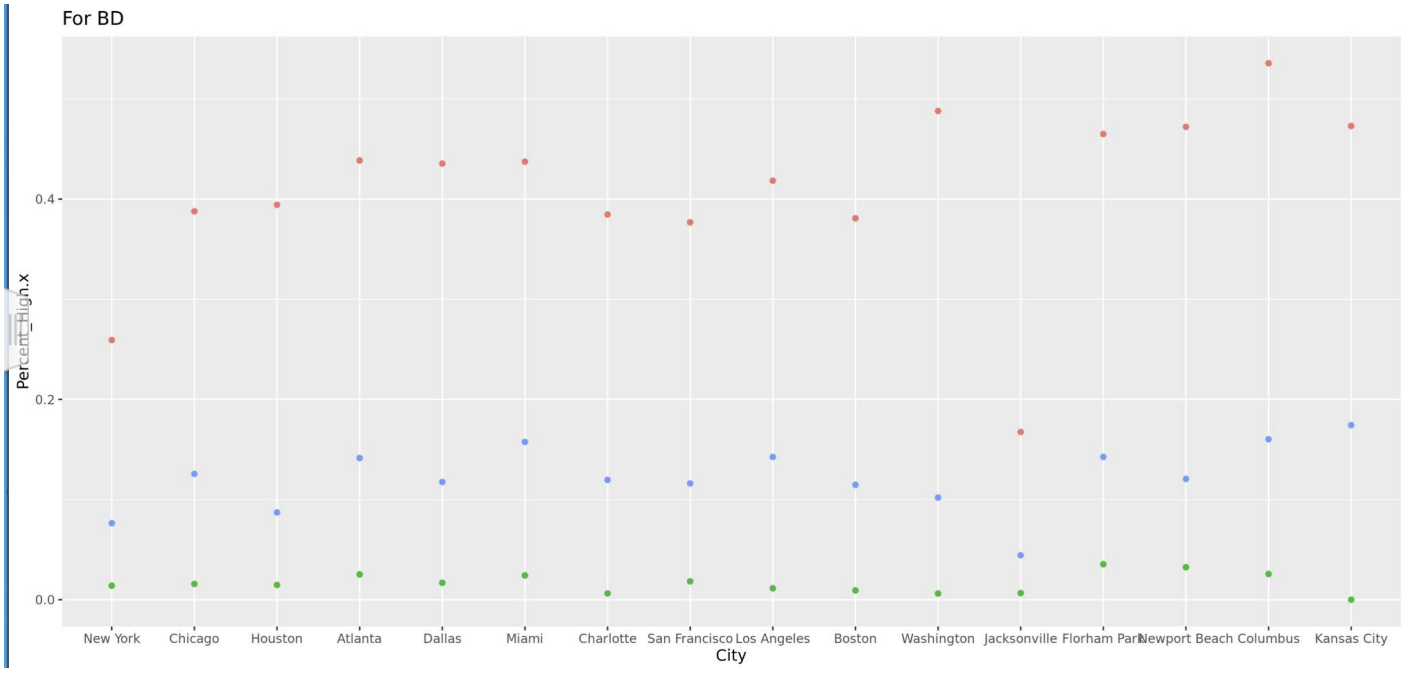
Var1	Freq
0	13534
1	47142
2	31692
3	19611
4	8984
5	3274
6	1251
7	978
8	370

(Var1 represents the number of changing titles, and Freq represents the frequency)

RIAs have more frequencies of 4,5,6,7,8 title changes than BDs.
BDs have more frequencies of 0,1,2,3 tile changes than RIAs.

10. **Are there specific branches/zip codes / MSAs that have more or more frequent title changes?**
    a. **Is it better to join, say ML, in NYC vs. Kansas City to progress in a career faster?**

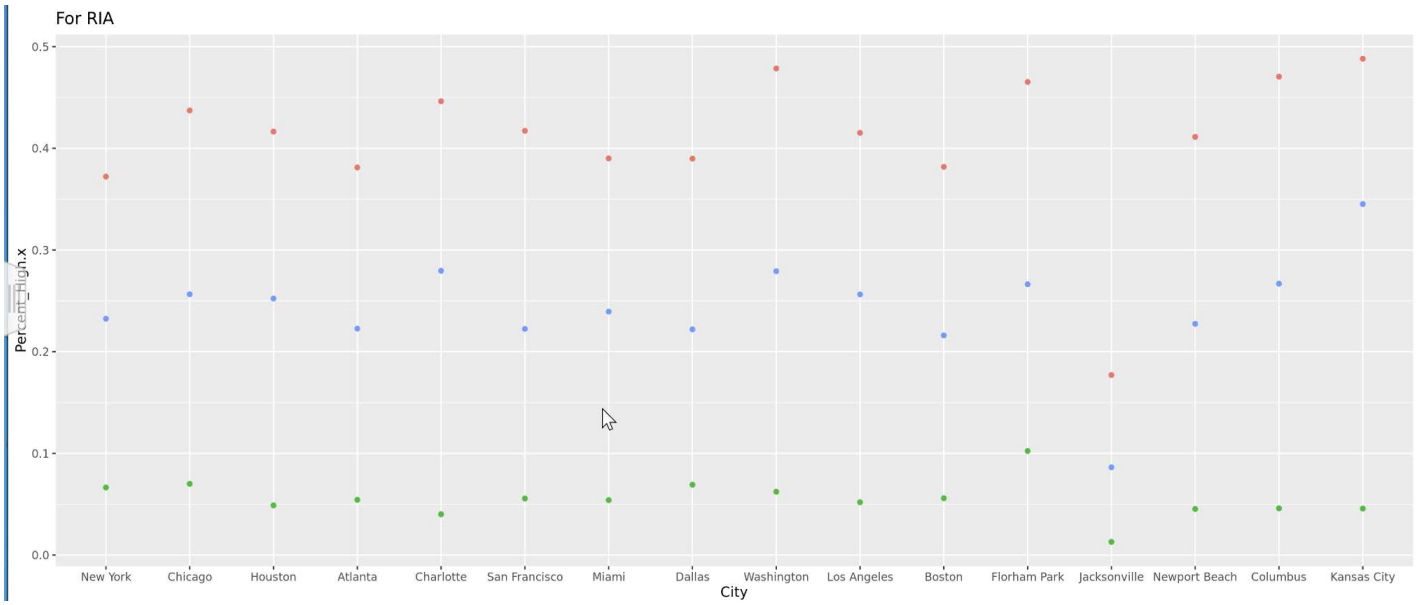
Ans:
For this question, I only consider ML.



The y variable is the percentage of the 'large' number of changing titles in each selected city.
The Upper line categorizes 3,4,5,6,7,8 as the 'large' number of changing titles in each selected city.
The Middle line categorizes 4,5,6,7,8 as the 'large' number of changing titles in each selected city.
The Lower line categorizes 5,6,7,8 as the 'large' number of changing titles in each selected city.

Ideas:
For ML(BD) case,
If considering 3,4,5,6,7,8 as the 'large' number of changing titles, I think it is better to join Kansas City than New York City. If considering 4,5,6,7,8 as the 'large' number of changing titles, I think it is better to join in Kansas City than in New York City. However, if considering only 5,6,7,8 as the 'large' number of changing titles, I think it is better to join New York City than Kansas City. The lower line indicates that there are almost no individuals who can have a number of title changes over 5. Compared to Kansas City, there are at least some individuals who can have a number of title changes over 5. That reflects that if the individual has a high degree of tolerance for transition changes, I think it would be better to join

Kansas City since the range is wider and the target is much easier to achieve. However, if the individual is highly ambitious, it would be a better choice to join New York City. Higher risk means higher returns. Generally, if the individual has a high degree of tolerance, he or she should join Columbus, Washington, Kansas City, Newport Beach, Florham Park, Atlanta, Miami, Dallas, and Los Angeles. However, if the individual has a low degree of tolerance, which is ambitious, he or she should join Florham Park, Newport Beach, Columbus, Atlanta, Miami, and San Francisco. Therefore, it is more interesting for me to join Florham Park, Newport Beach, Atlanta, and Miami, which are within the intersection of two sets.



The y variable is the percentage of the 'large' number of changing titles in each selected city.
The Upper line categorizes 3,4,5,6,7,8 as the 'large' number of changing titles in each selected city.
The Middle line categorizes 4,5,6,7,8 as the 'large' number of changing titles in each selected city.
The Lower line categorizes 5,6,7,8 as the 'large' number of changing titles in each selected city.

For ML(RIA) case,
If considering 3,4,5,6,7,8 as the 'large' number of changing titles, I think it is better to join Kansas City than New York City. If considering 4,5,6,7,8 as the 'large' number of changing titles, I think it is better to join in Kansas City than in New York City. However, if considering only 5,6,7,8 as the 'large' number of changing titles, I think it is better to join New York City than Kansas City. That reflects that if the individual has a high degree of tolerance for transition changes, I think it would be better to join Kansas City since the range is wider and the target is much easier to achieve. However, if the individual is highly ambitious, it would be a better choice to join New York City. Higher risk means higher returns. Generally, if the individual has a high degree of tolerance, he or she should join Kansas City, Columbus, Washington, Florham Park, Charlotte, Chicago, San Francisco, Houston, Los Angeles, and Newport Beach. However, if the individual has a low degree of tolerance, which is ambitious, he or she should join Florham Park, Chicago, New York, Dallas, Washington, San Francisco, and Boston. Therefore, it is more interesting for me to join Florham Park, Washington, Chicago, and San Francisco, which are within the intersection of two sets.

11. **Does tenure, age, and holding a designation (or multiple designations) drive career advancement (is it at least correlated)?**

Ans:

Since both BD and RIA data frames have no 'Age' column, I will use **'DateOfBirth_Year'** to investigate. **'n_change'** below means the number of title changes.

For BD,

```
> cor(BD_birs$DateOfBirth_Year,BD_birs$n_change)
[1] 0.02361946
```

The correlation coefficient of 'DateOfBirth_Year' and 'n_change' is 0.02361946.

For RIA,

```
> cor(RIA_birs$DateOfBirth_Year,RIA_birs$n_change)
[1] 0.02741686
```

The correlation coefficient between 'DateOfBirth_Year' and 'n_change' is 0.02741686. That may indicate that the correlation between the 'DateOfBirth_Year' and 'n_change' of RIA is slightly higher than that of BD.

Then, I would like to find the correlation between **'the mean of the number of holding designations'** and **'n_change'.**

For BD,

RepCRD	n_2013	n_2014	n_2015	n_2016	n_2017	n_2018	n_2019	n_2020	total	n_change
170	1	1	1	1	2	2	2	2	1.500	1
183	0	0	0	0	0	0	0	0	0.000	1
353	1	1	1	1	1	1	1	1	1.000	3
399	0	0	0	0	0	0	0	0	0.000	0
485	0	0	0	0	0	0	0	0	0.000	1
495	0	0	0	0	0	0	0	0	0.000	2
550	0	0	0	0	0	0	0	0	0.000	2
663	0	0	0	0	0	0	0	0	0.000	0
721	0	0	0	0	0	0	0	0	0.000	3
969	0	0	0	0	0	0	0	0	0.000	4
1185	0	0	0	0	0	0	0	0	0.000	0
1217	0	0	0	0	0	0	0	0	0.000	0
1275	0	0	0	0	0	0	0	0	0.000	0
1382	4	3	0	0	0	0	0	0	0.875	0

(part of the data, 'total' means the mean number of holding designations from 2013 to 2020. I use the mean here to diminish the effect of a large imbalance case on the whole dataset. In an imbalance case, some people hold many designations, like 5, in the first two periods but 0 in the rest periods, The sum, 10, is large but the average,1.25, is not very large.)

```
> cor(BD_n_design_tt$total, BD_n_design_tt$n_change)
[1] 0.1151532
```

The correlation coefficient of the mean number of holding designations from 2013 to 2020 and the number of transition title changes is 0.1151532.

For RIA,

RepCRD	n_2013	n_2014	n_2015	n_2016	n_2017	n_2018	n_2019	n_2020	total	n_change
170	1	1	1	1	2	2	2	2	1.500	1
183	0	0	0	0	0	0	0	0	0.000	1
353	1	1	1	1	1	1	1	1	1.000	3
399	0	0	0	0	0	0	0	0	0.000	0
485	0	0	0	0	0	0	0	0	0.000	1
495	0	0	0	0	0	0	0	0	0.000	2
550	0	0	0	0	0	0	0	0	0.000	2
663	0	0	0	0	0	0	0	0	0.000	0
721	0	0	0	0	0	0	0	0	0.000	3
969	0	0	0	0	0	0	0	0	0.000	4
1185	0	0	0	0	0	0	0	0	0.000	0
1217	0	0	0	0	0	0	0	0	0.000	0
1275	0	0	0	0	0	0	0	0	0.000	0
1382	4	3	0	0	0	0	0	0	0.875	0

to 15 of 344,120 entries, 11 total columns

(part of the data, 'total' means the mean number of holding designations from 2013 to 2020. I use the mean here to diminish the effect of a large imbalance case on the whole dataset. In an imbalance case, some people hold many designations, like 5, in the first two periods but 0 in the rest periods, The sum, 10, is large but the average,1.25, is not very large.)

```
> cor(RIA_n_design_tt$total, RIA_n_design_tt$n_change)
[1] 0.02644483
```

The correlation coefficient of the mean number of holding designations from 2013 to 2020 and the number of transition title changes is 0.02644483.

Then, I am interested in finding the correlation coefficient between **the number of changing titles** and **each product.**(CFA, CFP, CPA, ChFC, CLU).

For BD,

BD_CFA_corr	BD_CFP_corr	BD_CPA_corr	BD_ChFC_corr	BD_CLU_corr
0.04178468	0.1336933	0.02111809	0.02396636	0.06132473

(each column is the correlation coefficient between each product and 'n_changes')
For BD firms, CFP has a slightly higher correlation with the number of title changes than other products.)

For RIA,

RIA_CFA_corr	RIA_CFP_corr	RIA_CPA_corr	RIA_ChFC_corr	RIA_CLU_corr
0.1192187	0.07218551	0.04511257	-0.04479943	-0.01630074

(each column is the correlation coefficient between each product and 'n_changes')
For RIA firms, CFA has a slightly higher correlation with the number of title changes than other products. CLU has a *negative* correlation with the number of title changes.)

12. **Can career advancement be captured/identified through job titles (aka are job titles a valid proxy for career advancement)?**

Ans:

I think we can capture career advancement by using the transition matrix for top titles. If the entry value in some position is zero, say (1,2), then it means there is no probability of title 1 moving to title 2. If the value in (2,1) of the matrix is larger than 0, it means that title 2 can be able to move to title 1. The above may indicate that the grade of title 2 is lower than title 1. However, there is some exception that the individual may be degraded from title 2 to title 1, which may make the probability larger than 0. However, that would be a small probability event. Therefore, we indeed to investigate more about career advancement.

13. **Is the 'success likelihood' field an accurate measure of success over time? Does this field seem to correlate well with the title progression?**

Ans:

For this question, I will try to find the correlation between the success likelihood and the number of title changes. The success likelihood column contains two values, 'Higher' and 'Lower'. My definition of accuracy is the 'Higher' value in one period indicates the title change in the next period, and the 'Lower' value in one period indicates no title changes in the next period. I will use my definition of accuracy to examine whether the success likelihood field is an accurate measure of success over time.

(Since the success likelihood field only exists in the BD data frame, I will only consider BD.)

Period	pred_acclst
2013_2014	0.6392311
2014_2015	0.6437849
2015_2016	0.5494091
2016_2017	0.5391068
2017_2018	0.5126597
2018_2019	0.4880350
2019_2020	0.5030496

(the pred_acclst is the accuracy for each period)

Then, I define the overall accuracy score as the mean of the pred_acclst column.

The value of the overall accuracy score is 0.5536109, which is higher than 0.50 and lower than 0.60.

Therefore, I think the success likelihood is not an accurate measure of success over time.