Diversity and Inclusion Report

This report uses randomly generated data and does not reflect the situation in a perticular company.

Dr. Philippe De Brouwer

20 April, 2021

Table of Contents

The Particular Example Behind this Demo Report

Overview

Diversity

Inclusion

Conclusions

Appendices

The Particular Example Behind this Demo Report

The Particular Example Behind this Demo Report

The data used for this report is randomly generated with the following characteristics:

- team size: 400
- percentage of females: 0.35
- average male's salary / average female's salary: 1.035
- no other biases are built in (so any other observations stem from random generation of data. This can be seen as "despite no bias against citizenship by the manager, some pay-gaps will be different from one." This means that being unbiased is not necessarily the same as having equal outcomes.)

Equitable outcomes are not the same as equal outcomes!

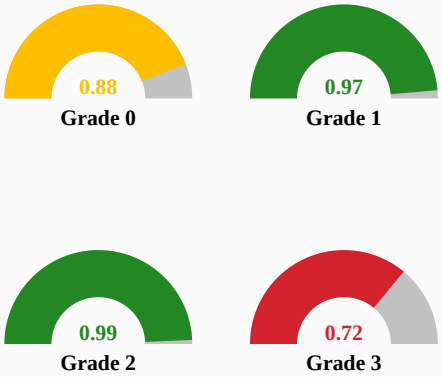
Overview

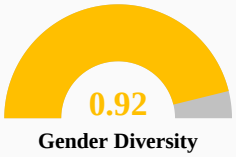
Findings (in order of importance)

Nbr	Area	Finding	Suggestion
1	Gender	Where we can calculate the paygap between females and non-females, we find that the females generally earn less in similar roles and similar grades.	Check the gender-paygap table and identify the grade/role combinations where an the paygap has most stars. Check if the salary differences are justified.
2	Age	The team is predominantly younger than the surrounding population (Poland).	Consider hiring older people to balance. Focus on retention.
3	Gender	The diversity is good in grade 1 and 2, but under par in grade 3	Consider if females have barriers to apply to grade 3 jobs and remove the barriers.
4	Gender	Males in Grade 2 seem to have been promoted faster.	Understand unconscious bias, coach everyone (and specially females), work on trust.

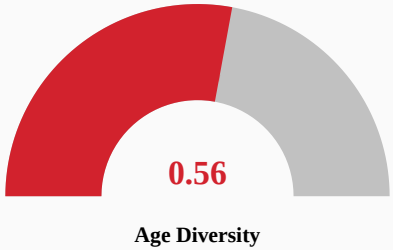
Dashboard

The diversity is good in grade 1 and 2, but under par in grade 3

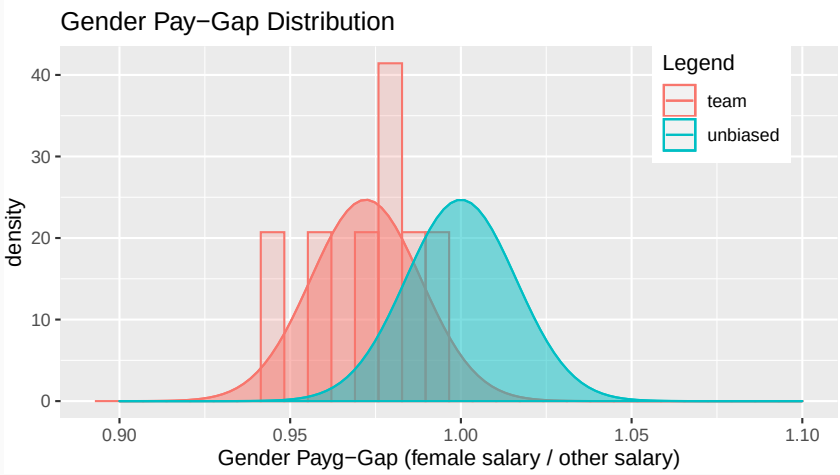




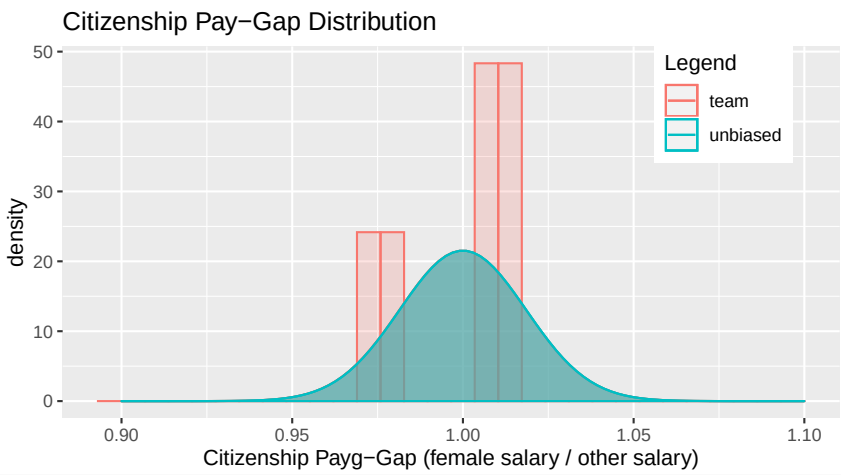
The team is predominantly younger than the surrounding population (Poland).



Where we can calculate the paygap between females and non−females, we find that the females generally earn less in similar roles and similar grades.

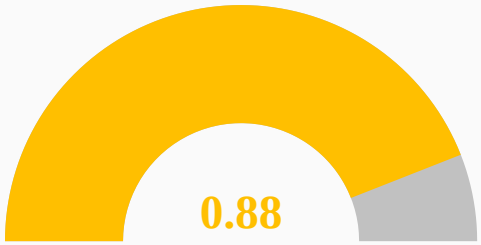


No Bias detetected for Citizenship (in salary) −− both team and unbiased distribution are virtually the same

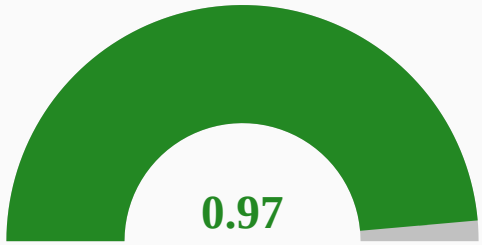


Diversity

Gender diversity per grade

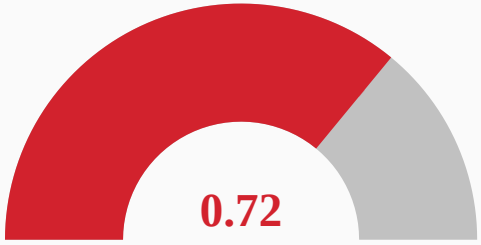


Grade 0



Grade 1

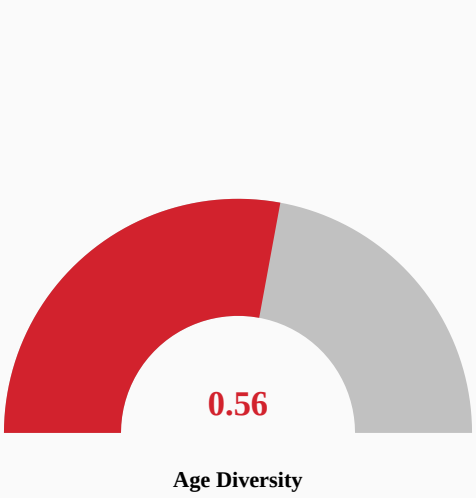
Grade 2



Grade 3

Figure 1: The diversity of the team with respect to gender per grade.

Age diversity



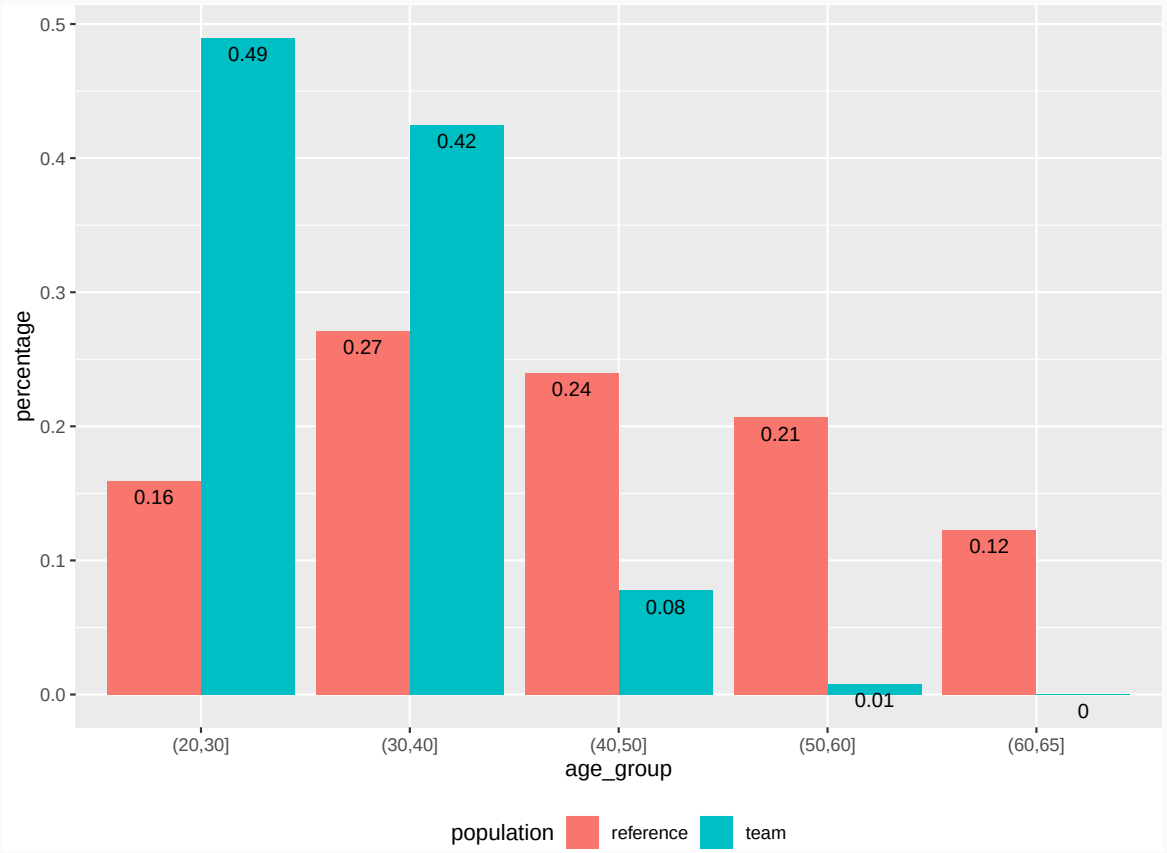


Figure 2: The diversity of the team with respect to age, assuming the age distribution of the country as reference.

Diversity in nationalities (1/2)

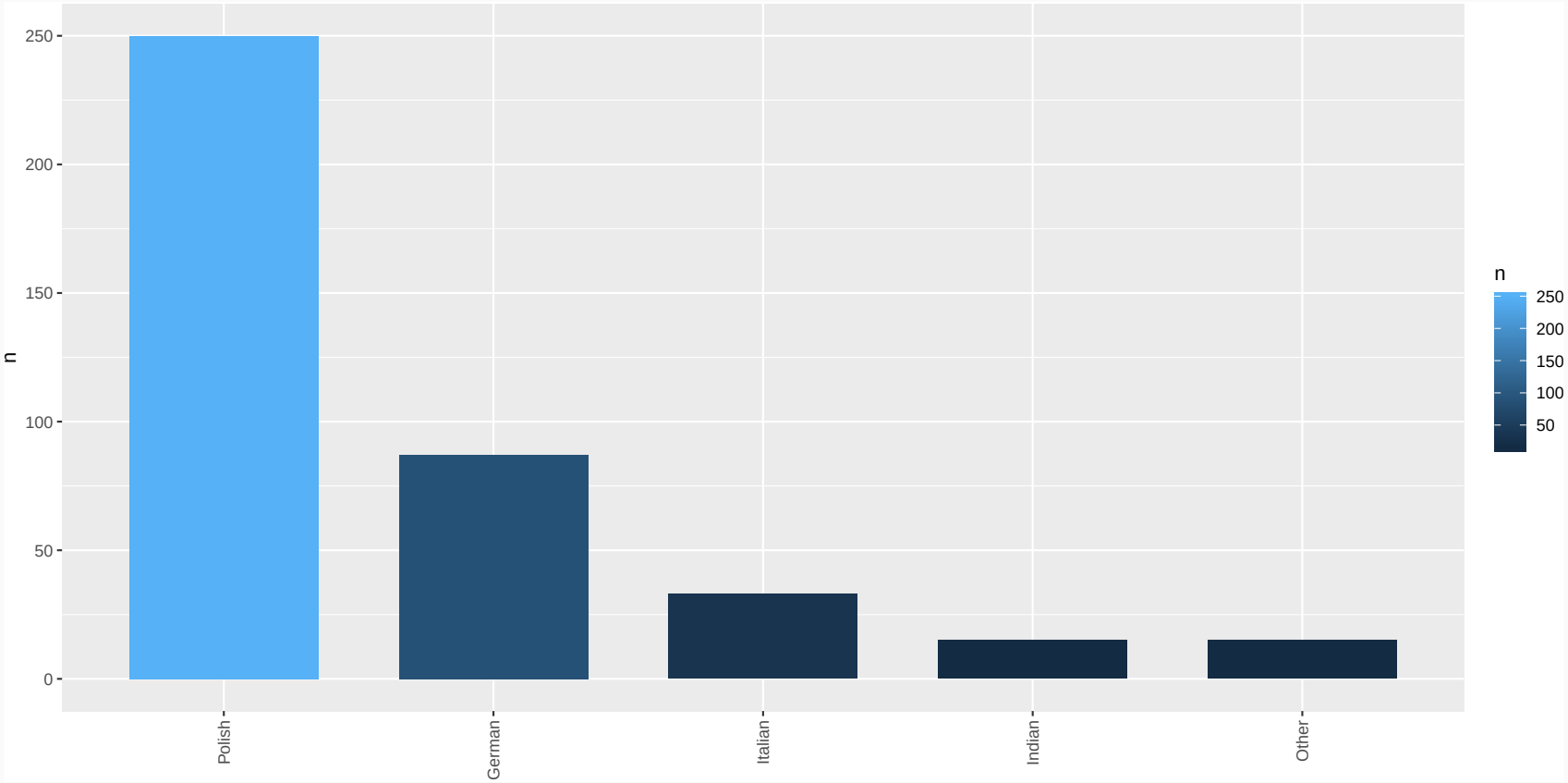


Figure 3: The barplot for the nationalities in the team over all grades.

Diversity in nationalities (2/2)

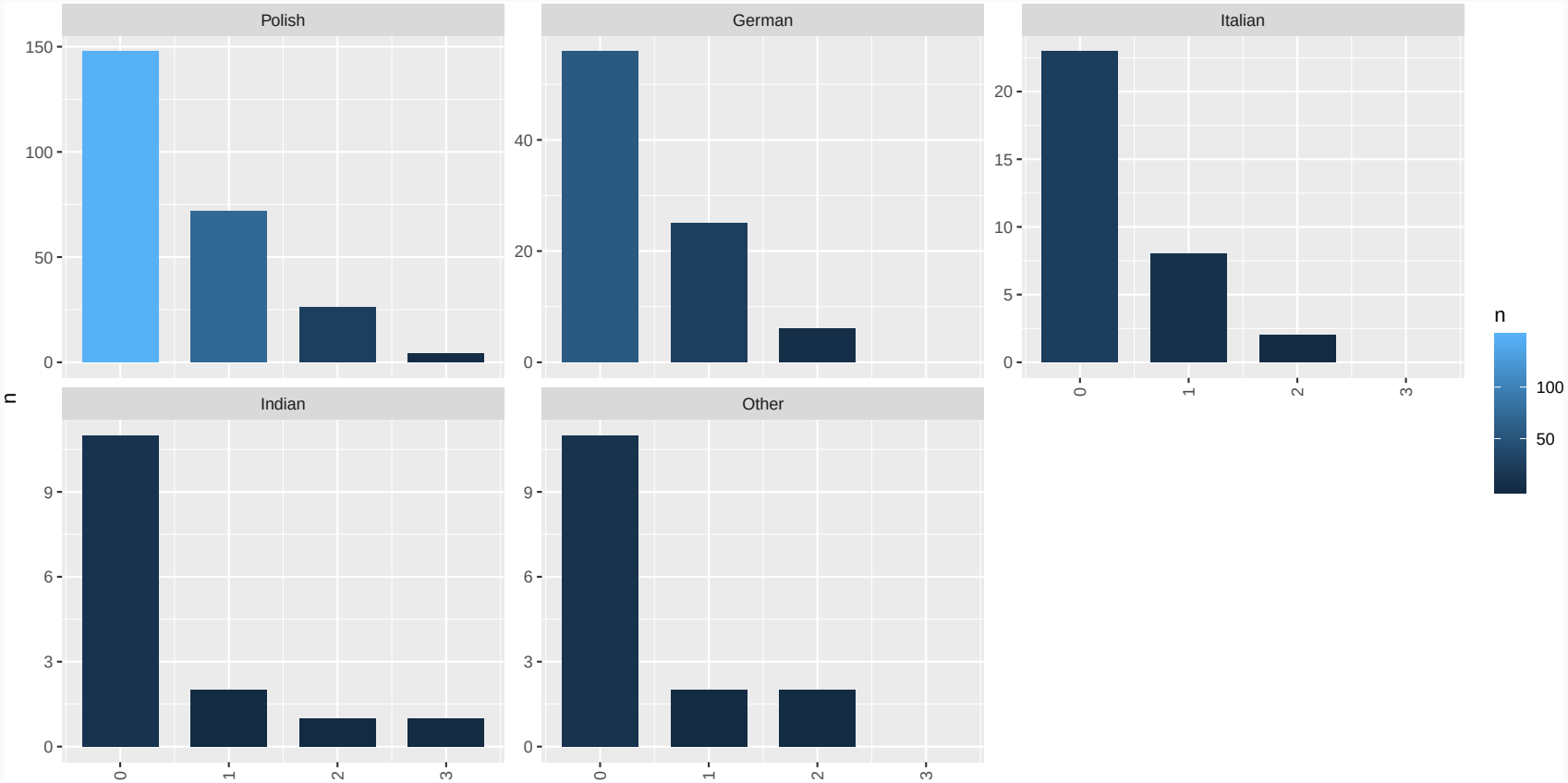


Figure 4: The breakdown of each grade per nationalitiy.

Inclusion

The Gender PayGap

Table 1: The paygap for gender (in terms of salary) as a ratio, along with the confidence level that this paygap is significant alongside the control variable age.

grade	jobID	sal_F	sal_oths	n_F	n_oths	med_age_F	med_age_o	paygap	p-value	conf.
0	sales	3,902	4,133	51	105	28.0	29.0	0.944	0.008647	**
2	sales	17,971	18,737	12	16	34.5	35.5	0.959	0.000670	***
3	sales	38,154	39,326	1	3	34.0	39.0	0.970	0.500000	
1	analytics	8,500	8,703	17	24	31.0	32.0	0.977	0.092868	.
2	analytics	18,022	18,443	4	5	37.0	36.0	0.977	0.063492	.
0	analytics	4,177	4,229	24	69	27.0	29.0	0.988	0.396839	
1	sales	8,625	8,712	27	41	32.0	31.0	0.990	0.349614	
3	analytics	NA	38,825	0	1	NA	43.0	NA	NA	NA

The Citizenship PayGap

Table 2: The paygap for citizenship (in terms of salary) as a ratio, along with the confidence level that this paygap is significant alongside the control variable age.

grade	jobID	sal_Polis	sal_oths	n_Polish	n_oths	med_age_P	med_age_o	paygap	p-value	conf.
2	sales	18,244	18,700	19	9	34.0	36.0	0.976	0.307855	
1	analytics	8,560	8,761	26	15	32.5	30.0	0.977	0.694702	
0	analytics	4,227	4,193	56	37	28.5	28.0	1.008	0.275245	
0	sales	4,078	4,042	92	64	27.0	29.0	1.009	0.664176	
2	analytics	18,207	17,947	7	2	35.0	39.0	1.014	0.888889	
1	sales	8,702	8,569	46	22	32.5	30.5	1.016	0.553660	
3	sales	39,035	NA	4	0	37.0	NA	NA	NA	NA
3	analytics	NA	38,825	0	1	NA	43.0	NA	NA	NA

The Age Paygap

Table 3: The paygap for age (in terms of salary) as a ratio, along with the confidence level that this paygap is significant alongside the control variable age.

grade	jobID	sal_L	sal_H	n_L	n_H	med_age_L	med_age_H	paygap	p-value	conf.
1	analytics	8,465	8,741	19	22	26.0	35.0	0.968	0.056311	.
2	sales	18,130	18,584	14	14	32.0	40.5	0.976	0.163552	
3	sales	38,740	39,115	2	2	34.5	43.5	0.990	0.666667	
0	analytics	4,200	4,221	40	53	25.0	32.0	0.995	0.568433	
1	sales	8,676	8,609	34	34	29.5	36.0	1.008	0.692023	
2	analytics	18,270	18,112	4	5	32.0	42.0	1.009	0.412698	
0	sales	4,106	4,042	72	84	24.0	32.0	1.016	0.340726	
3	analytics	NA	38,825	0	1	NA	43.0	NA	NA	NA

Time in firm paygap

Table 4: The paygap for tenure firm (in terms of salary) as a ratio, along with the confidence level that this paygap is significant alongside the control variable age.

grade	jobID	sal_L	sal_H	n_L	n_H	med_age_L	med_age_H	paygap	p-value	conf.
1	sales	8,566	8,854	34	34	33.0	31.0	0.968	0.078323	.
0	sales	4,040	4,125	78	78	28.0	28.0	0.979	0.614746	
2	sales	18,228	18,535	14	14	35.5	35.0	0.983	0.874287	
3	sales	38,740	39,115	2	2	34.5	43.5	0.990	0.666667	
0	analytics	4,237	4,207	46	47	28.0	28.0	1.007	0.865754	
2	analytics	18,270	18,112	4	5	32.0	42.0	1.009	0.555556	
1	analytics	8,652	8,495	20	21	33.5	30.0	1.018	0.705278	
3	analytics	NA	38,825	0	1	NA	43.0	NA	NA	NA

Job Changes per Year per Gender

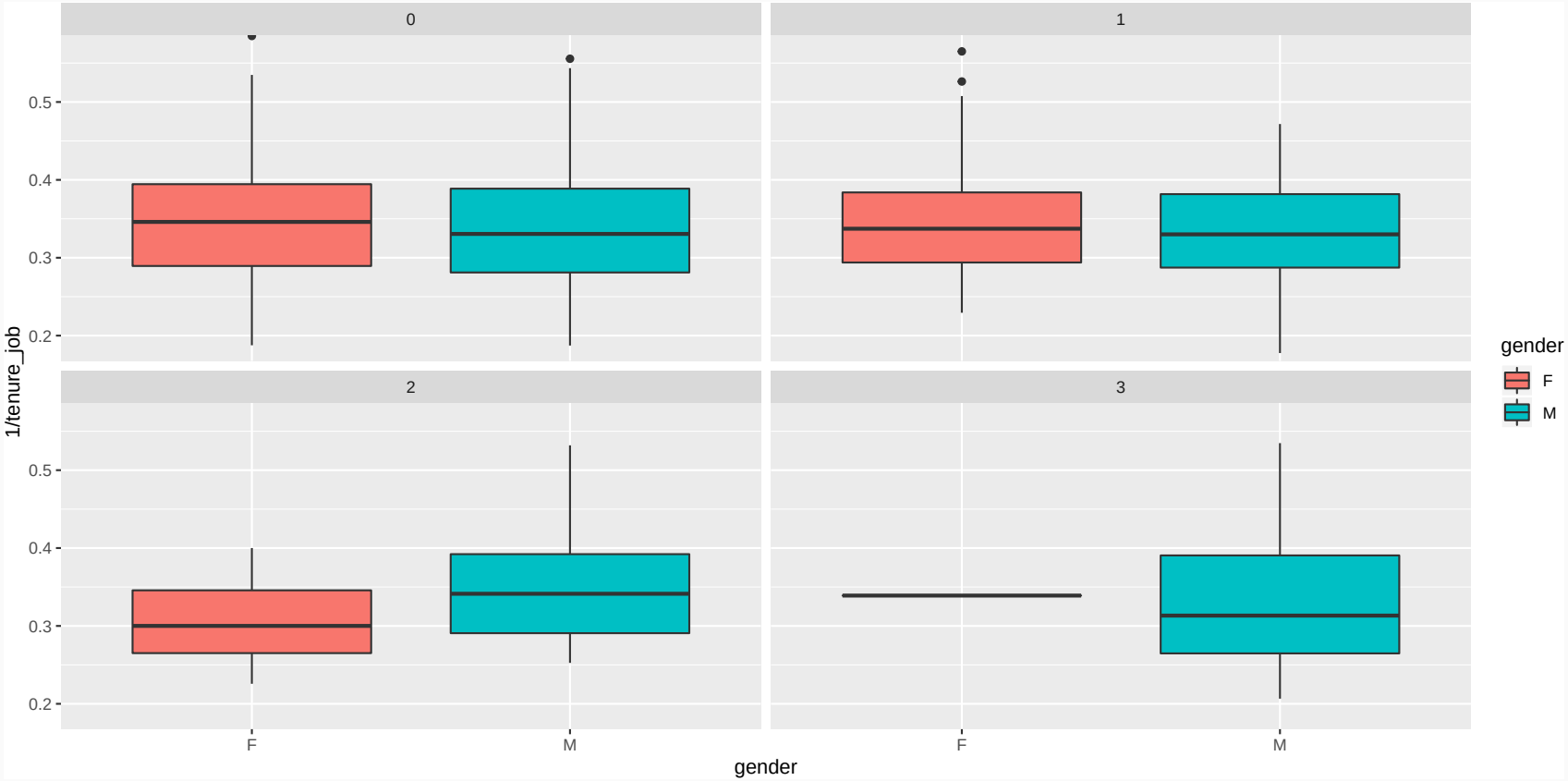


Figure 5: Job changes per year indicate mobility and risk taking. They are a good indication for promotion (see Figure 6).

Promotions per Year per Gender

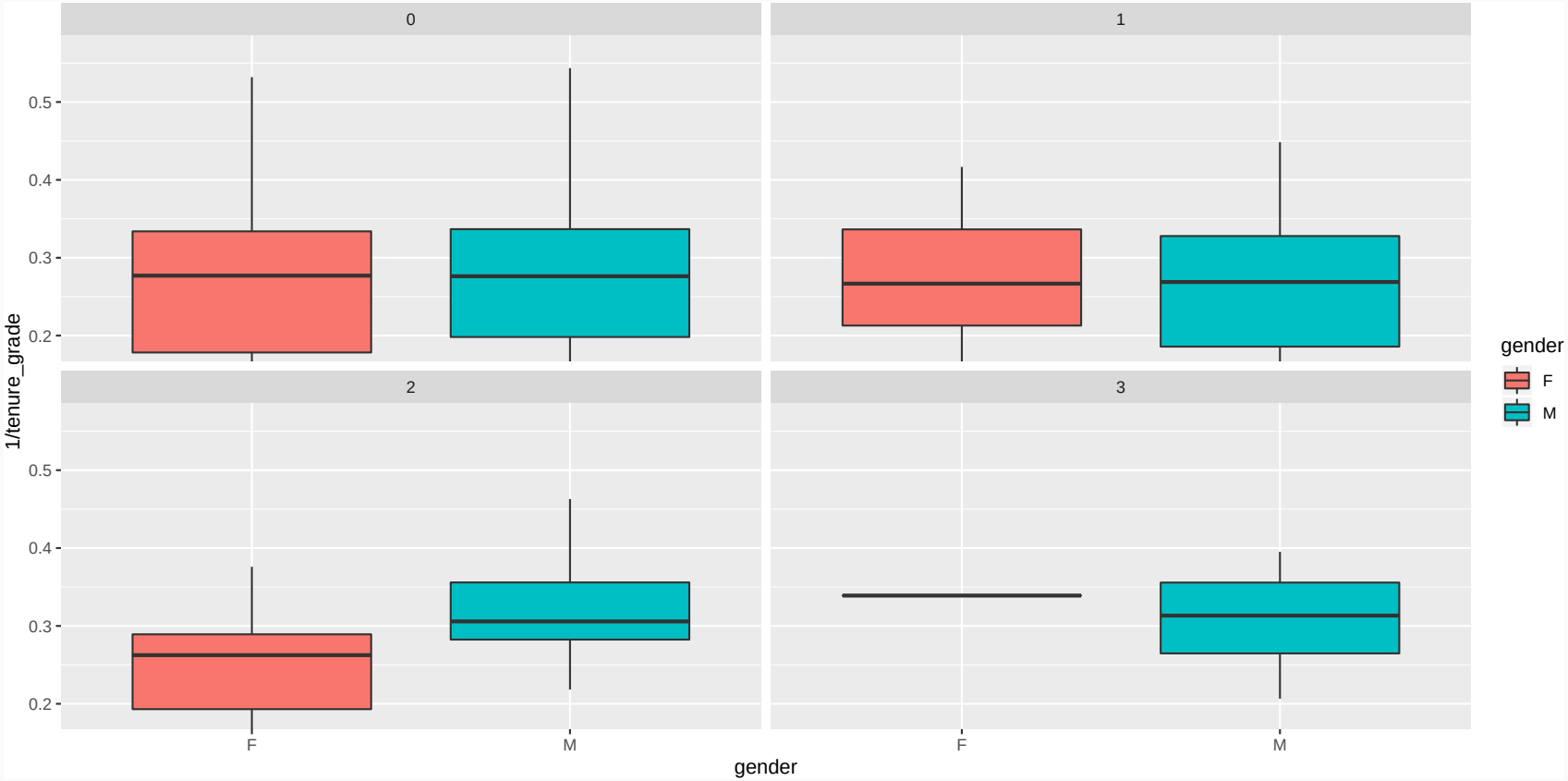


Figure 6: The number of promotions per year can show if a gender is more probable to be promoted.

Conclusions

Conclusions

Nbr	Suggestion
0	Learn more by reading e.g. "The Essentials of Diversification & Inclusion", Dabrowska (2019)
1	Check the gender-paygap table and identify the grade/role combinations where an the paygap has most stars. Check if the salary differences are justified.
2	Consider hiring older people to balance. Focus on retention.
3	Consider if females have barriers to apply to grade 3 jobs and remove the barriers.
4	Understand unconscious bias, coach everyone (and specially females), work on trust.

Appendices

Legend Paygap

- paygap = the ratio of median salaries of one group divided by the median of the salaries of the other group
- 'NA' = numbers are too small, please look at individuals;
- nothing = no bias detectable;
- '.' = maybe there is some bias, but the numbers are low, check individuals;
- '*' = you should check for bias;
- '**' = bias is probably there;
- '***' = most certainly there is bias

So, there will be more stars if the probability of a bias is higher: this can be due to a higher bias and/or due to a larger sample size.

Legend: Paygap Column Headers

- `grade` = the salary grade as used in the company
- `jobID` = a unique identifier of the job category (can be abbreviated)
- `sal_F` = the median salary of the females (F)
- `sal_oth` = the median salary of the other groups (non F). The tool is open to use more than one gender.
- `age_F` = the median age of the females (or `age_Pol` could be the median age of the team members with Polish citizenship)
- `age_oth` = the median age of the other groups take together (e.g. the median age of non females)
- `paygap` = the ratio of median salary earned by the selected group (e.g. females) divided by the median of the other people. If this is lower than 1, then median female earns less than the median non-female.
- `conf.` = the confidence level that this paygap is significant.

The Diversity Index (1/2)

We express diversity as a number between zero and one. Our calculation is based on De Brouwer (2020) and more in particular section 36.3.1 "The Business Case: a Diversity Dashboard' '. Details can be found in the book. The method is:

- The diversity is 0 if only one of the groups is present, and is 1 if both groups are equitably present.
- This calculation is similar to the established concept of entropy in physics.
- More than two categories can be used (e.g. one is not limited to two genders)
- We calibrate the probabilities so that they show maximum entropy (or diversity) for the percentages that naturally occur (see next slide).

The Diversity Index (2/2)

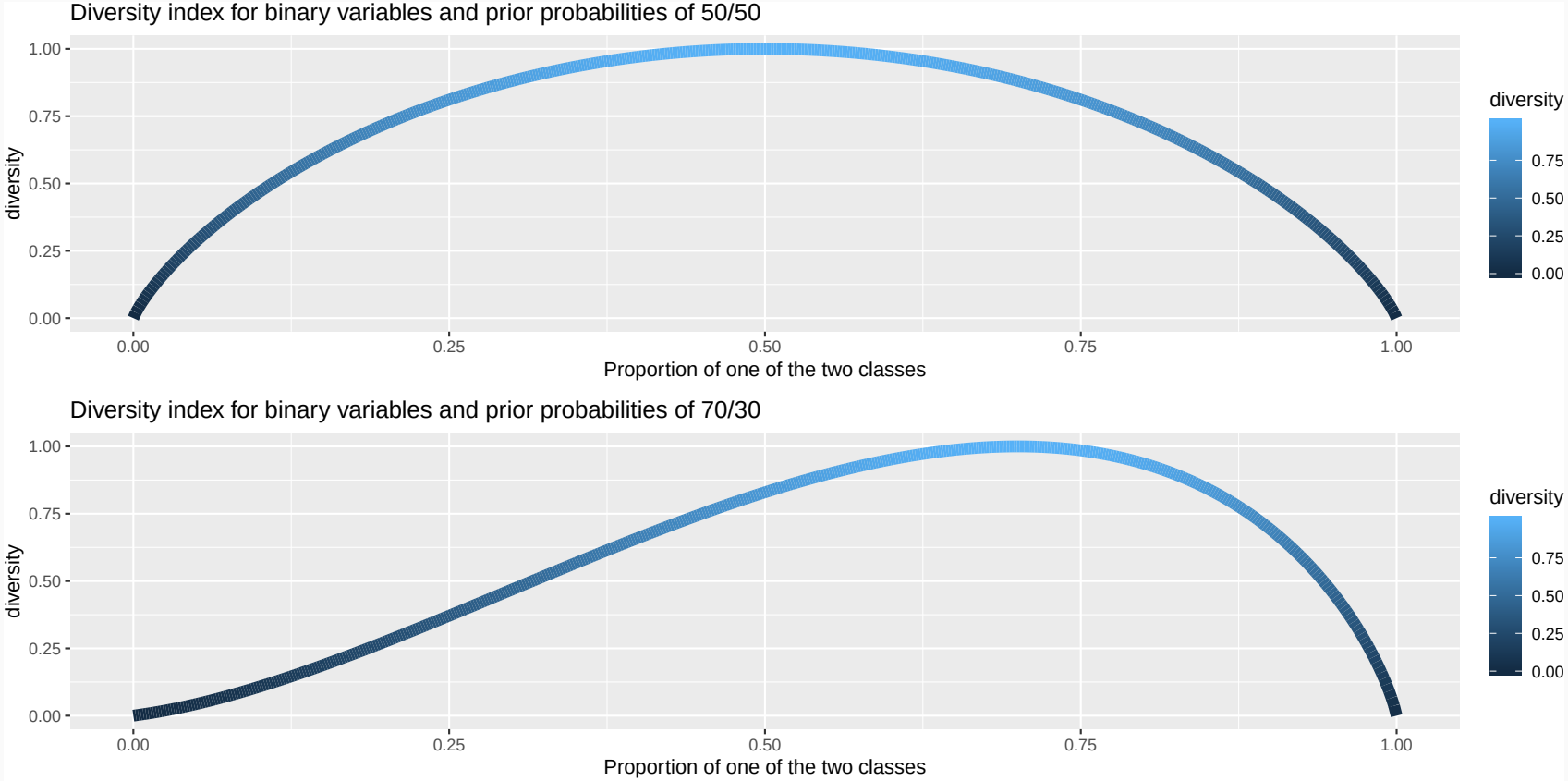


Figure 7: The diversity index illustrated for the case where there are only two possible classes, and where the prior priorities are respectively 50/50 (top) and 70/30 (bottom). The index reaches a maximum at a distribution equal to the prior probabilities.

The confidence level and p-value

The p-value is the probability that we make a mistake by assuming that there is no paygap.

It is calculated by splitting the data on a variable in binary factors (e.g. Females and others) and then checking how likely it is that a random person from the first group earns less than a random person from the second group. This is done by a method known as Mann-Whitney U test: see Wikipedia[1]

[1] The Mann–Whitney U test (aka. Mann–Whitney–Wilcoxon (MWW), Wilcoxon rank-sum test, or Wilcoxon–Mann–Whitney test) is a nonparametric test of the null hypothesis that, for randomly selected values X and Y from two populations, the probability of X being greater than Y is equal to the probability of Y being greater than X. If we assume that the distributions are symmetric, it boils down to a test that the medians are different.

Another view on the PayGap

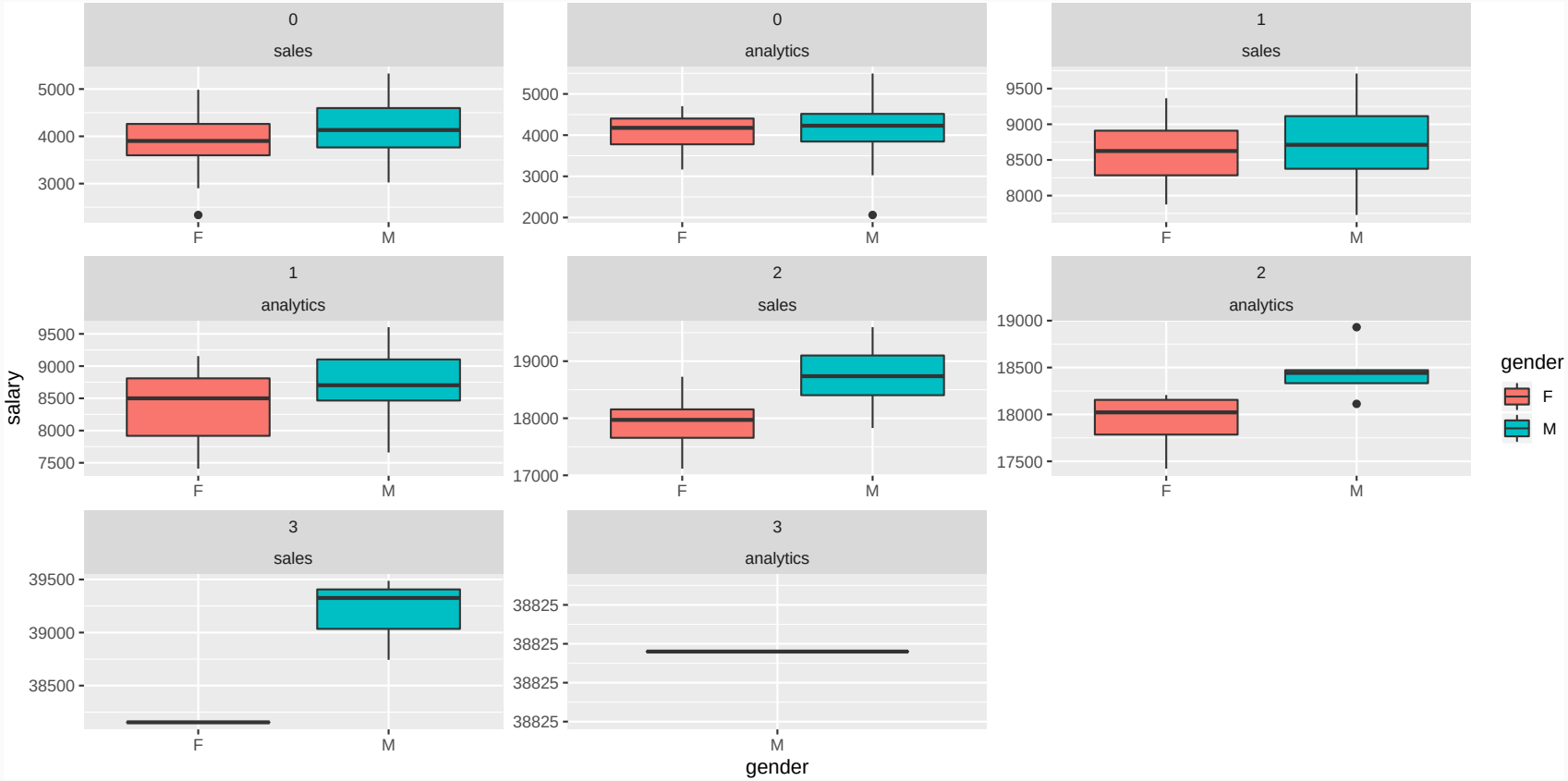


Figure 8: Boxplots for each grade (over all job categories) per gender. This another view of the same data as in Table 1.

Bibliography

De Brouwer, Philippe J.S. 2020. *The Big r-Book: From Data Science to Learning Machines and Big Data*. New York: John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781119632757.

Zaroda-Dabrowska, Anna, and Tomasz Dabrowski. 2019. *The Essentials of Diversity & Inclusion Management*. Krakow: AT Wydawnictwo.