EXPLORING LEADERSHIP STYLES FOR INNOVATION

30 November 2020

Submitted By
SATYAM VATTS (100368408),
AVNEET KAUR (100368395),
RADHIKA MAINI (100340257),
KIRAN VIKRAM ANGA (100368125)

PURPOSE

This report describes the analysis of Australian Public Service Employee Survey conducted from May 11 to 12 June in the year 2015. This analysis is aimed to explore different leadership styles that are considered significant for creativity & innovation in workplace in Australia. APS Leadership survey dataset containing response to various questions relating to employee wellbeing, productivity, impressions of managers, workgroup and agency is analysed for:

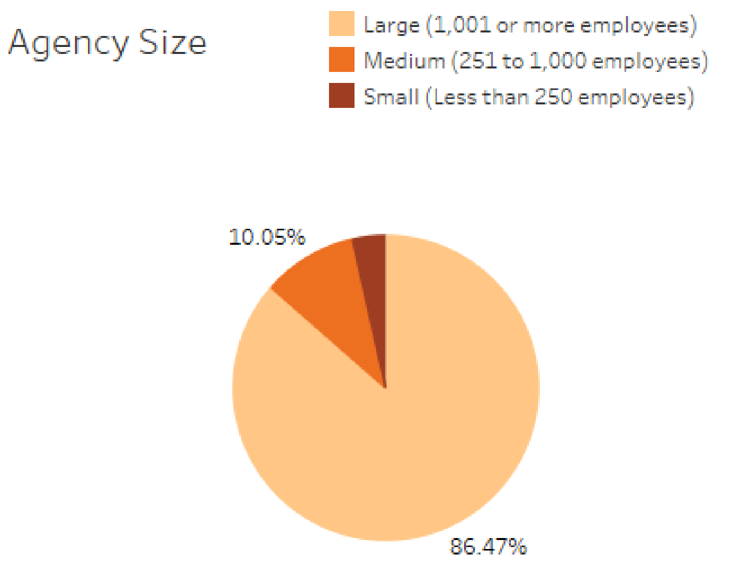
- Identifying the participants in Survey
- Providing Demographic profile of Participants
- Reporting and Treating Missing Data
- Analysing data type and Normality assumptions
- Detecting and Dealing with Outliers
- Performing Descriptive Analysis
- Identifying difference in Satisfaction
- Performing Exploratory Factor Analysis
- Identifying characteristics of Major Leaders in APS
- Establishing differences of opinions on immediate supervisor & workgroup between males and females
- Comparing Wellbeing & Productivity impressions of Employees from 2014 to 2015
- Investigating Wellbeing and Engagement Index
- Recommendations
- Conclusion

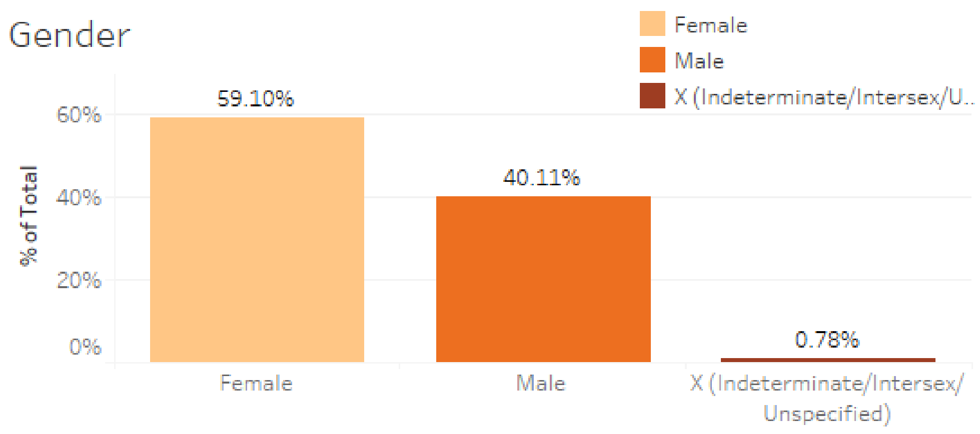
DEMOGRAPHIC ANALYSIS

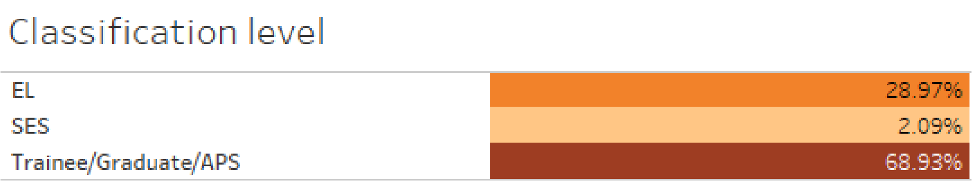
The participants/respondents of the APS survey are all the available and eligible Australian public Service employees. The APS survey for the year 2015 has a response rate of 66%.

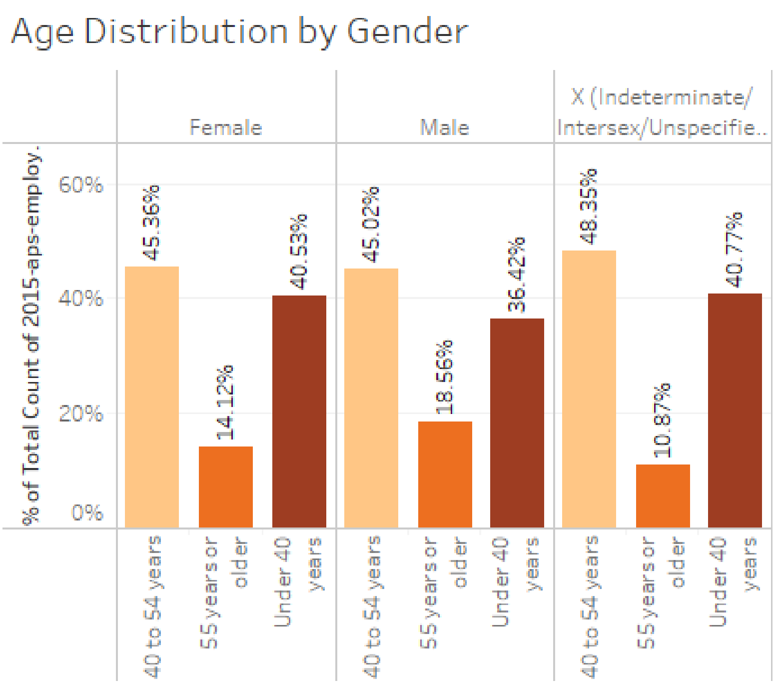
The demographic analysis represents:

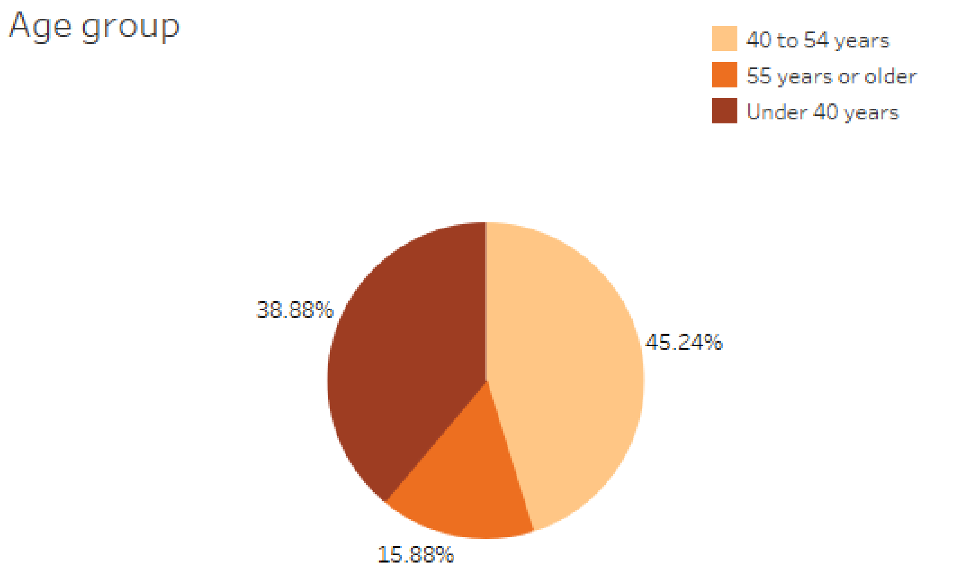
- 59% of the sample population were female and 40% were males while less than 1% were intersex or indeterminate.
- Majority of the respondents (45%) were aged between 40 to 55 years or less than 40 years (39%).
- 86.5 % of the population worked in large agencies with more than 1000 employees
- 68% had classification level as Trainee or Graduates.











HANDLING MISSING DATA

The dataset considered for analysis is a sampled version of actual responses. The dataset consists of 45250 rows and 261 columns. The demographics of data above described the population distribution. But the dataset has a lot of missing data. Below table describes the amount of missing data for selected variables of an analysis.

Variable	Missing	Variable	Missing	Variable	Missing	Variable	Missing	Variable	Missing
AC1	0	***q20e***	690	***q24p***	2212	***q37b***	4017	***q58e***	3562
AS2	0	***q20f***	687	***q30***	2121	***q37c***	4029	***q58f***	3647
q1	326	***q20g***	680	***q31***	9015	***q37d***	4018	***q58g***	3619
q2	21	***q20h***	715	***q32***	2174	***q37e***	4073	***q58h***	3610
q7@	0	***q20i***	728	***q33***	2204	***q37f***	4097	***q58i***	3571
q8	158	***q21a***	940	***q34a***	3345	***q37g***	4178	***q58j***	3543
q18a	503	***q21b***	951	***q34b***	4667	***q37h***	4238	***q59***	3652
q18b	506	***q21c***	981	***q35a***	2287	***q37i***	4254	***q60***	3373
q18c	529	***q21d***	1043	***q35b***	2283	***q37j***	4355	***q64c***	5074
q18d	549	***q21e***	1053	***q35c***	2319	***q37k***	4286	***q67c***	3608
q18e	537	***q21f***	1133	***q35d***	2291	***q37l***	4303	***q67d***	3652
q18f	545	***q21g***	1098	***q35e***	2295	***q53b***	3092	***q67e***	3630
q19	502	***q21h***	1064	***q35f***	2283	***q53c***	3139		
q20a	621	***q21i***	1080	***q35g***	2346	***q58a***	3417		
q20b	648	***q21j***	1112	***q36a***	3750	***q58b***	3484		
q20c	708	***q24m***	2302	***q36b***	23967	***q58c***	3496		
q20d	669	***q24o***	2244	***q37a***	3964	***q58d***	3576		

It can be observed that there is a lot of missing data particularly in questions q36b with 24k missing values. Also, the missing data is randomly scattered throughout the dataset.

The missing data is MCAR i.e., missing completely at random. The reason being the respondents have not replied to certain questions which results in null values. Since these null values do not relate to observed data and have no relation with non-missing observations, these are regarded as MCAR.

Although there are techniques such as Estimation Maximization to impute missing data, we have not imputed the missing values due to following aspects:

- The responses are individual preferences mentioned by respondents. Based on other variables, it is not logical to predict these responses. It can lead to biased dataset eventually distorting our analysis.
- The dataset is huge even after removing the observations containing any null values. Below is the data structure after removing missing values:
  - Number of Observations: 13065
  - Number of Variables selected: 84
- Also, the distribution of participants based on gender, classification, age group etc. has not changed after removing the null data. It can be observed from the below visuals if compared to actual dataset.

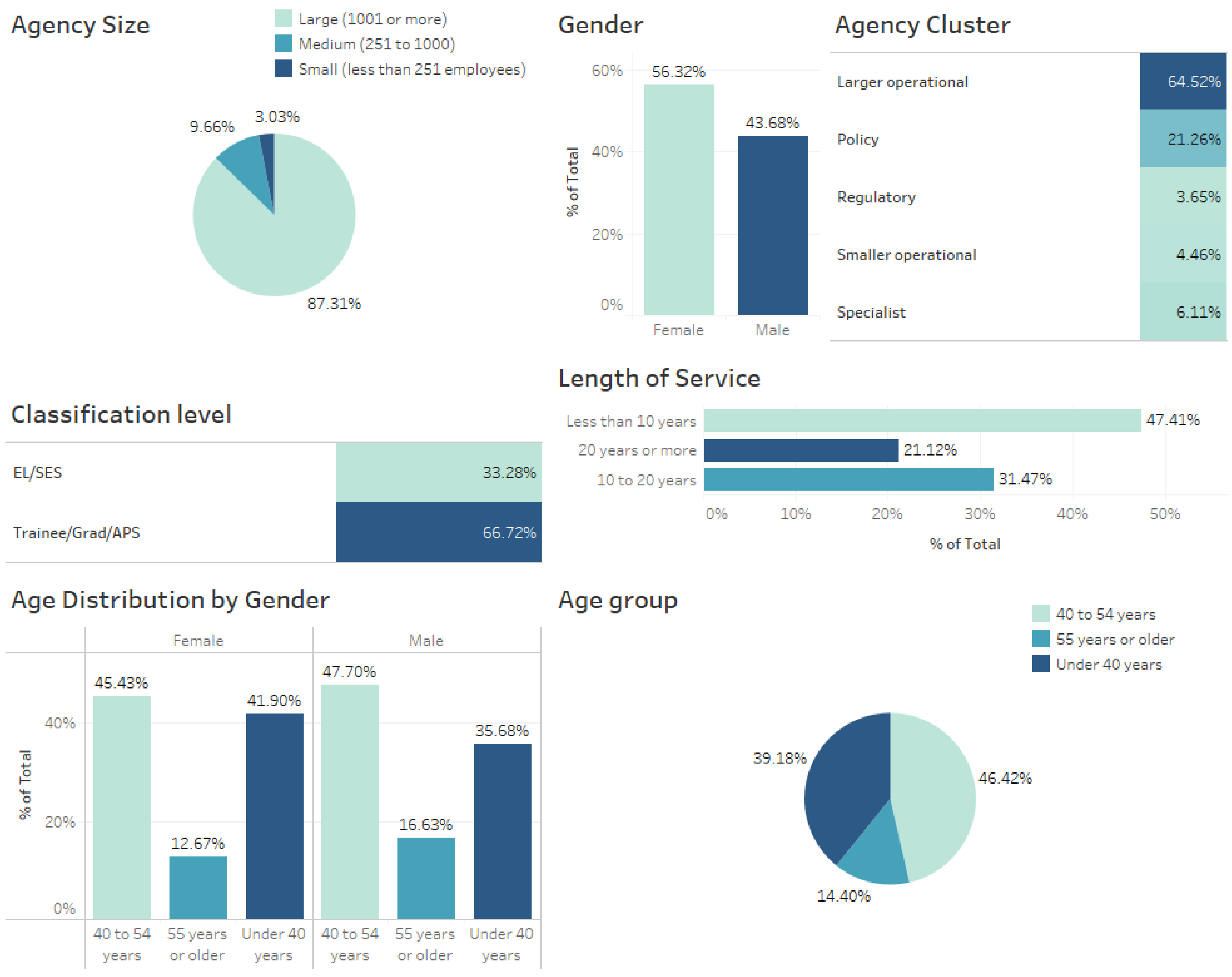
After cleaning the dataset, the final dataset ready for further analysis consists of 13K responses consisting of following information:

- Description of Participants: AC1, AS2, q1, q2, q7, q8
- General Impressions about Immediate Workgroup: q18(a to f), q19
- General Impressions about Immediate Supervisor: q20(a to i)
- General Impressions about Immediate Senior Leadership: q21(a to j)
- General Impressions about Managers in Agency: q24(m, o, p)
- Wellbeing: q30, q31, q32, q33, q34(a,b),q35(a to g), q36(a,b), q37(a to l)
- Performance Management: q53(b,c)
- Leadership: q58(a to j), q59
- Innovation: q60
- Leaves & Code of Conduct: q64c, q67(c,d,e)

DEMOGRAPHIC ANALYSIS OF THE CLEANED DATASET

The demographic factors undertaken in this report are Agency Size, cluster, Gender and Age distriution, classification level, and Total length of service. The demographic analysis represents:

- 56% of the respondants were females aged between 40 to 54 years (45%). On the other hand, males constituted 44 % of the population aged between 40 to 54 years(47.7%).
- Majority (46.4 %) of the respondants belonged to the age group of 40 – 54 years followed by those who were aged less than 40 years (39%).
- 87% worked in Large agencies and 64.5 % belonged to operational agency cluster.
- Approximately half of the population (47%) had total length of service less than 10 years.
- Two third of the population (approx 67%) were either Trainees or graduates whereas the other one third had classification level as SES or EL.



NORMALITY & OUTLIERS

Analytically, the normality of the data can be determined by calculating the skewness of the data. The rule of thumb says, +/- 2.58 of skewness value can be approximately normal. Skewness is a metric to describe how skewed the data is from the mean. From the above table, the values of skewness range from 0.6 to -0.6, which are all under the threshold limit. Similarly, kurtosis is a measure of the sharpness of the peak in the data distribution. These values range from 0.1 to -1.05, which are in the recommended range for the data to be normal.

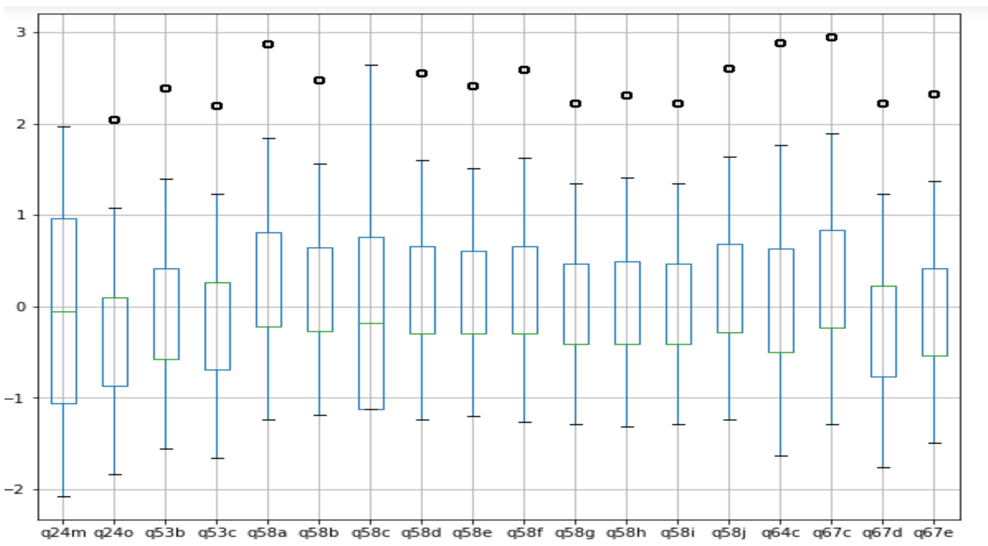
There is a persuasive argument against considering the normality of data for a Likert scale data, and in a way, it is true as per our knowledge. Having said that, the Likert scale data can very well be denoted with numerical values making it easy to bring out the descriptive analysis like means, standard deviation. However, the tricky part is to interpret these results. In general, normality is checked or considered for some tests (Parametric tests), which assumes or expects that the data is normal. So, if we decide what tests/analysis we are going to perform on our data, it is easy to consider the normality of the data. If we are not performing any test (non-Parametric tests) which requires the data to be normal, we can very well not consider the normality of the data.

As stated with evidence by Dr. Geoff Norman, one of the world's leaders in medical education research methodology, that parametric tests can be used with ordinal data, like from Likert scales, and they are more powerful than non-parametric tests. Even when statistical assumptions—such as a normal distribution of data—are violated, parametric tests tend to give "the right answer." Educators and researchers utilize z-score standardization of several Likert-type items. It is recommended for less concrete aspects like the one we are trying to analyse which are trainee motivation, employee satisfaction, and leadership—where a single survey item is unlikely to be capable of fully capturing the concept being assessed. Thus, parametric tests are sufficiently robust to yield unbiased answers that are close to "the truth" when analysing Likert scale responses.

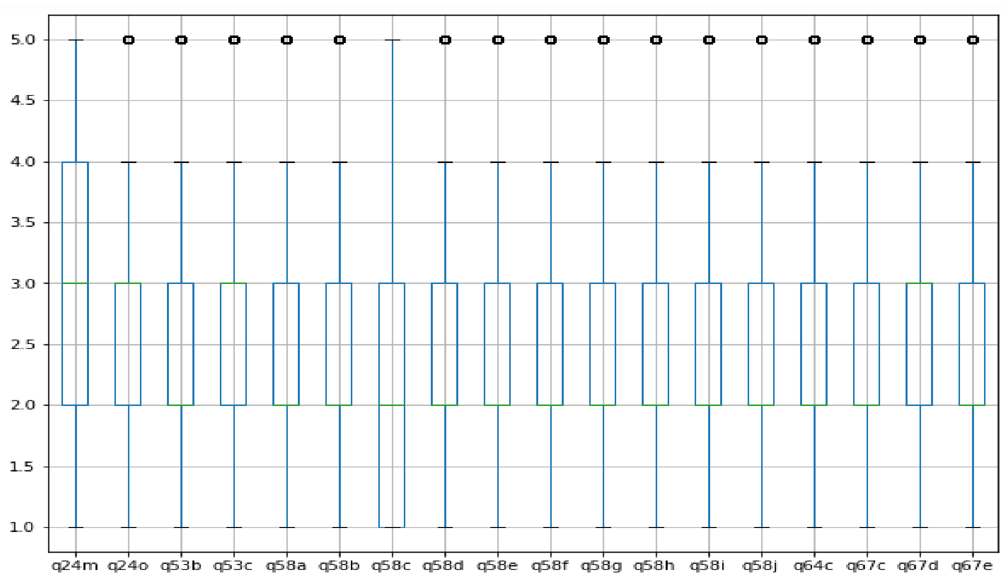
Outliers are determined by calculating the absolute value of z-scores greater than 3.29 for each column. Below is the maximum and minimum Z-Score standardized values of the selected questions.

Questions	Max- Z Score	Min- Z score
q24m	1.969623513	-2.075671232
q24o	2.048266851	-1.838672604
q53b	2.381680319	-1.556264234
q53c	2.19313324	-1.65281633
q58a	2.86649232	-1.23983395
q58b	2.480552009	-1.18721757
q58c	2.647911467	-1.11773165
q58d	2.548292585	-1.235375639
q58e	2.412751962	-1.196740997
q58f	2.585735938	-1.259426971
q58g	2.21688346	-1.288118343
q58h	2.314452278	-1.318432808
q58i	2.224899572	-1.285130363
q58j	2.599222974	-1.240755333
q64c	2.888307804	-1.627562528
q67c	2.951246145	-1.290094839
q67d	2.219553079	-1.757060583
q67e	2.329473491	-1.492404657

Below is the boxplot of z-score standardized values of 18 variables. It can be noticed from the box plot given below that all the values are less than 3, hence there are no outliers in the 18 variables selected for EFA.



Although the boxplot shows some points as the outliers for all questions except 24m, but the below plot of actual values/data shows that the values shown as outliers in the above boxplot are valid/accepted responses within the range of 1 to 5.



DESCRIPTIVE ANALYSIS

Mean, Standard Deviation, Skewness and Kurtosis

Question	Variable Description	mean	std	skew	kurtosis
q20b	My supervisor provides me with regular and constructive feedback	2.355236	1.091423	0.713342	-0.17719
q20d	My supervisor works effectively with people from diverse backgrounds	2.009002	0.868729	1.020657	1.547675
q20e	My supervisor is committed to workplace safety	1.967531	0.784109	0.86141	1.556411
q20f	My supervisor is accepting of people from diverse backgrounds	1.88505	0.787339	1.018157	1.947086
q20g	My supervisor treats people with respect	1.968685	0.947379	1.220556	1.549959
q24m	Managers in Agency Encourage innovation	3.05243	0.988803	0.120197	-0.35805
q24o	Managers in Agency Value their Employees	2.892155	1.029087	0.301296	-0.32799
q53b	Supervisor provides with clear and consistent performance expectations	2.580788	1.015758	0.71516	-0.06279
q53c	Supervisor provides with a clear understanding of how performance is assessed and measured	2.71902	1.040055	0.510637	-0.39758
q58a	Achieves results	2.207731	0.974107	0.901024	0.625164
q58b	Cultivates productive working relationships	2.294757	1.090581	0.814811	0.05258
q58c	Exemplifies personal drive and integrity	2.187294	1.062236	0.905549	0.337446
q58d	Shapes strategic thinking	2.306008	1.057175	0.744029	0.080134
q58e	Communicates with influence	2.326215	1.108189	0.772462	-0.06143
q58f	Sets direction	2.310142	1.040268	0.761358	0.170169
q58g	Motivates people	2.470034	1.141226	0.614525	-0.3451
q58h	Encourages innovation	2.451665	1.101053	0.616668	-0.21721
q58i	Develops people	2.464524	1.139591	0.638294	-0.30884
q58j	Is open to continued self learning	2.292461	1.041673	0.80151	0.327977
q64c	Immediate supervisor actively manages employees' sick/carer's leave	2.441638	0.885765	0.720884	0.617496
q67c	Immediate supervisor takes responsibility when a problem is identified	2.216686	0.943098	0.953707	0.899802
q67d	Senior leaders in agency take responsibility when a problem is identified	2.767394	1.005881	0.399032	-0.18329
q67e	People in agency are encouraged to speak up when they identify a serious policy or delivery risk	2.561959	1.046606	0.692843	-0.06657

The descriptive analysis was based on the values of mean and standard deviation. The mean was useful in describing the average opinion of the respondents to obtain the overall picture of the respondents' perceptions regarding each variable. So, from the above analysis, we concluded that the mean value of the leadership of innovation was ranging from level 2 to 3, but most values were lying close to 2 and a little bit higher.
Looking at q24m we analysed that some of the managers were encouraging towards the innovation as mean is 3 which denotes 'Somewhat'. Similar observation can be made for 'valuing their employees' as the average is near 3.
Looking at the mean responses of the q53 b and c we analysed that the respondents agreed to the fact that their supervisors provided correct responses to the respondents about their performance which meant that if there was a need to improve then the respondents could easily improve as there was a clear-cut communication between the supervisors and the respondents.
Looking at the responses of q58 it was clear that the supervisors were exhibiting the officer like qualities, and they were quite motivating when it came to their employees. Most of the employees were satisfied with the leadership style which provided them support and gave them the opportunity to improve.
Looking at the responses of q64 it was clear that supervisors strongly cared about their employees' health and if there was a need of healthcare leave it was granted to the employees.
Finally, q67 reflects that Managers and Supervisors take responsibility of the problems they identify thus proving as inspiration for subordinates.

ANOVA – Test of differences in opinions of Managers and Subordinates

One Way Analysis Variance (ANOVA) was conducted on two levels of Classification - Executive Level/SES (Managers) and Trainee/Graduates (Subordinates) to compare their means and therefore to determine the differences in the responses of the two groups.

Null Hypothesis (H_0): The means of the two groups are equal

Alternate Hypothesis (H_a): The means of the two groups are not equal

Below is the table depicting F- Ratio representing ratio of symmetric variance to unsymmetric variance along with its p - value and mean differences.

If p – value is less than 0.05, we reject null hypothesis and conclude there is significant difference in opinions of the two groups – Managers and Subordinates. Otherwise, when p – value is greater than 0.05, we accept the null hypothesis and conclude opinions of both the groups doesn't differ.

Variable	F - Ratio	p-value	Mean EL/SES	Mean Trainee/Grad/APS	Mean Diff	Interpretation
q20b	0.711523	0.398954	2.366659	2.349527	0.017132	Opinion regarding providing regular & constructive feedback by supervisors doesn't differ between managers and subordinates
q20d	29.4318	5.90E-08	1.950589	2.038195	0.087606	Significant difference of opinions exists between managers and subordinates about supervisor effectiveness in working with people from diverse background
q20e	27.90998	1.29E-07	1.916186	1.993192	0.077006	Significant difference of opinions exists between managers and subordinates about supervisor's commitment to workplace safety
q20f	45.89945	1.30E-11	1.818979	1.918071	0.099091	Significant difference of opinions exists between managers and subordinates about supervisors' acceptance of people from diverse background

q20g	16.77287	4.24E-05	1.920573	1.99273	0.072158	Significant difference of opinions exists between managers and subordinates about receiving respect from supervisor
q24m	2.198653845	0.138155443	3.034268629	3.061489044	0.027220415	Opinion regarding encouragement on innovation doesn't differ between managers and subordinates
q24o	49.7153618	1.87E-12	2.802437902	2.936904898	0.134466996	Significant difference of opinions exists between managers and subordinates about valuing the employees in their agency.
q53b	2.23202585	0.13520066	2.599586017	2.571412183	0.028173833	Opinion regarding clear performance expectations from supervisors doesn't differ between managers and subordinates
q53c	24.79726304	6.45E-07	2.783118675	2.687048296	0.096070379	Significant difference of opinions exists between managers and subordinates about getting informed of performance assessment.
q58a	85.72734874	2.38E-20	2.096366145	2.263278651	0.166912506	Significant difference of opinions exists between managers and subordinates about leadership skills of immediate supervisors.
q58b	55.47769007	1.00E-13	2.194342226	2.344843409	0.150501183	
q58c	109.5507838	1.55E-25	2.050137994	2.255707239	0.205569244	
q58d	126.9643145	2.58E-29	2.159153634	2.379258919	0.220105285	
q58e	62.93149158	2.31E-15	2.217571297	2.380406103	0.162834806	
q58f	56.56249878	5.80E-14	2.213431463	2.358380177	0.144948714	
q58g	22.5593914	2.06E-06	2.402943882	2.50349891	0.100555028	
q58h	57.27240323	4.05E-14	2.348666053	2.503040037	0.154373983	
q58i	36.78243419	1.36E-09	2.379024839	2.507169898	0.128145059	
q58j	44.01644019	3.38E-11	2.20699172	2.335092348	0.128100628	
q64c	0.841255769	0.3590548	2.451701932	2.436618103	0.015083829	Significant difference of opinions exists between managers and subordinates about getting sick leaves from managers.
q67c	30.70445212	3.06E-08	2.152023919	2.248938855	0.096914936	Significant difference of opinions exists between managers and subordinates about taking responsibilities of a problem by supervisor.
q67d	0.839096885	0.3596721	2.755979761	2.773087071	0.01710731	Opinion regarding taking responsibilities of a problem by senior leaders doesn't differ between managers and subordinates
q67e	7.046166182	0.007953135	2.596366145	2.544797522	0.051568623	Significant difference of opinions exists between managers and subordinates about speaking up on serious policy risks

EXPLORATORY FACTOR ANALYSIS

Exploratory Factor Analysis (EFA) is performed on the above selected survey questions to extract latent factors that describe leadership styles in Australian Public Service agencies. Since we are performing exploratory analysis, we have tested two models to check the presence of leadership factors.

Assumption Check

The assumptions that are required to be tested before proceeding with EFA are Normality & Sampling Adequacy. The following tests are performed in R to verify the assumptions:

- Bartlett's Test of Sphericity: Multivariate Normality & Correlation
- Kaiser-Meyer-Olkin Test: Sampling Adequacy

Model 1: Supervisor, Senior Leaders and Agency Impressions

The first model was developed using all the questions listed in previous section. The following are the results of assumption check to verify validity of data for conduction EFA.

```
$chisq
[1] 2033.597

$p.value
[1] 3.552442e-275

$df
[1] 253
```

Bartlett's Test

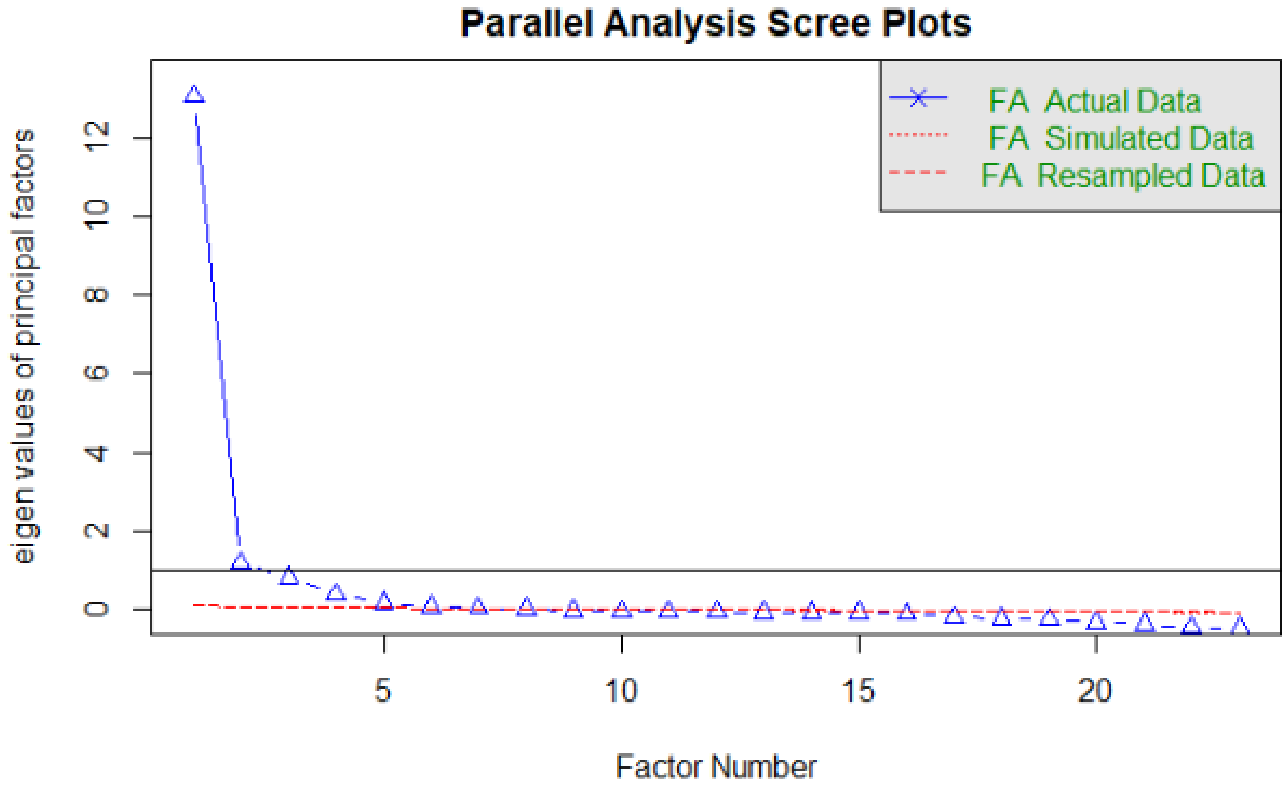
```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = correlation)
Overall MSA =  0.97
MSA for each item =
q24m q24o q53b q53c q58a q58b q58c q58d q58e q58f q58g q58h q58i q58j q64c
0.92 0.93 0.94 0.94 0.98 0.98 0.98 0.98 0.98 0.98 0.98 0.98 0.98 0.99 0.99
q67c q67d q67e q20b q20d q20e q20f q20g
0.98 0.91 0.94 0.99 0.95 0.98 0.93 0.97
```

KMO Test

It can be observed that:

- Bartlett's Test results in p-value of 3.5 e-275 which is very less than 0.05. Thus, the variables have significant correlation and therefore are suitable for structure detection. It also helps us to select rotation for factor analysis model development. Since high correlation exists, we will select oblimin.
- KMO Test: The measure of Sampling Adequacy is 0.97 which is high and indicate high proportion of variance in variables that are caused by underlying factors. A factor analysis is thus, useful for such a dataset.

Below is the Parallel Analysis scree plot for this model describing eigen values of principal factors.

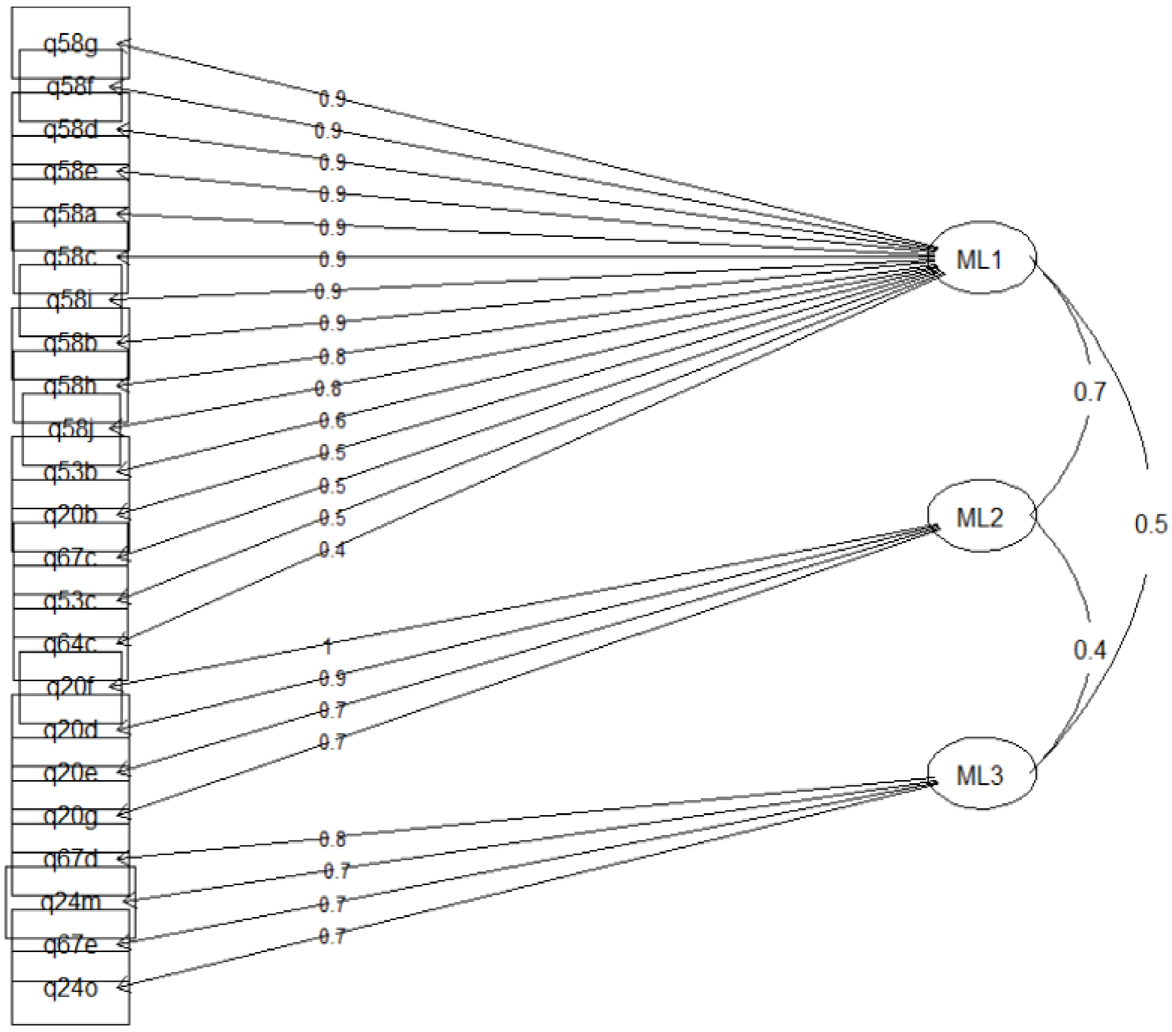


	Eigen_Value
1	13.06901745
2	1.19579896
3	0.81140952
4	0.38015596
5	0.17454900
6	0.07420134
7	0.05297784
8	0.01743205
9	-0.02912921
10	-0.04327470
11	-0.04697908
12	-0.05540238
13	-0.07132368
14	-0.07460107
15	-0.08536978
16	-0.10190931
17	-0.19069635
18	-0.22543564
19	-0.26042778
20	-0.32550106
21	-0.37289666
22	-0.46392559
23	-0.48552894

It can be observed that based on New Kaiser Criterion, 3 factors have eigen values greater than 0.7. Thus, we are considering 3 factors for this model for conducting factor analysis.

The model developed with 3 factors results in the following mapping of questions to factors.

Supervisor: Motivates people
Supervisor: Sets direction
Supervisor: Shapes strategic thinking
Supervisor: Communicates with influence
Supervisor: Achieves results
Supervisor: Exemplifies personal drive and integrity
Supervisor: Develops people
Supervisor: Cultivates productive working relationships
Managers: Encourage innovation
Supervisor: Is open to continued self-learning
Supervisor: Provides with suitable performance expectations
Supervisor: Regular and constructive feedback
Supervisor: Takes responsibility when a problem is identified
Supervisor: Provides clear understanding of performance Assessment
Supervisor: Manages employees' sick/carer's leave
Supervisor: Accepting of people from diverse backgrounds
Supervisor: Works with people from diverse backgrounds
Supervisor: Committed to workplace safety
Supervisor: Treats people with respect
Senior leaders: Take responsibility when a problem is identified
Agency: Encourages to speak up in case of serious policy or delivery risk
Supervisor: Encourages innovation
Managers: Value their Employees

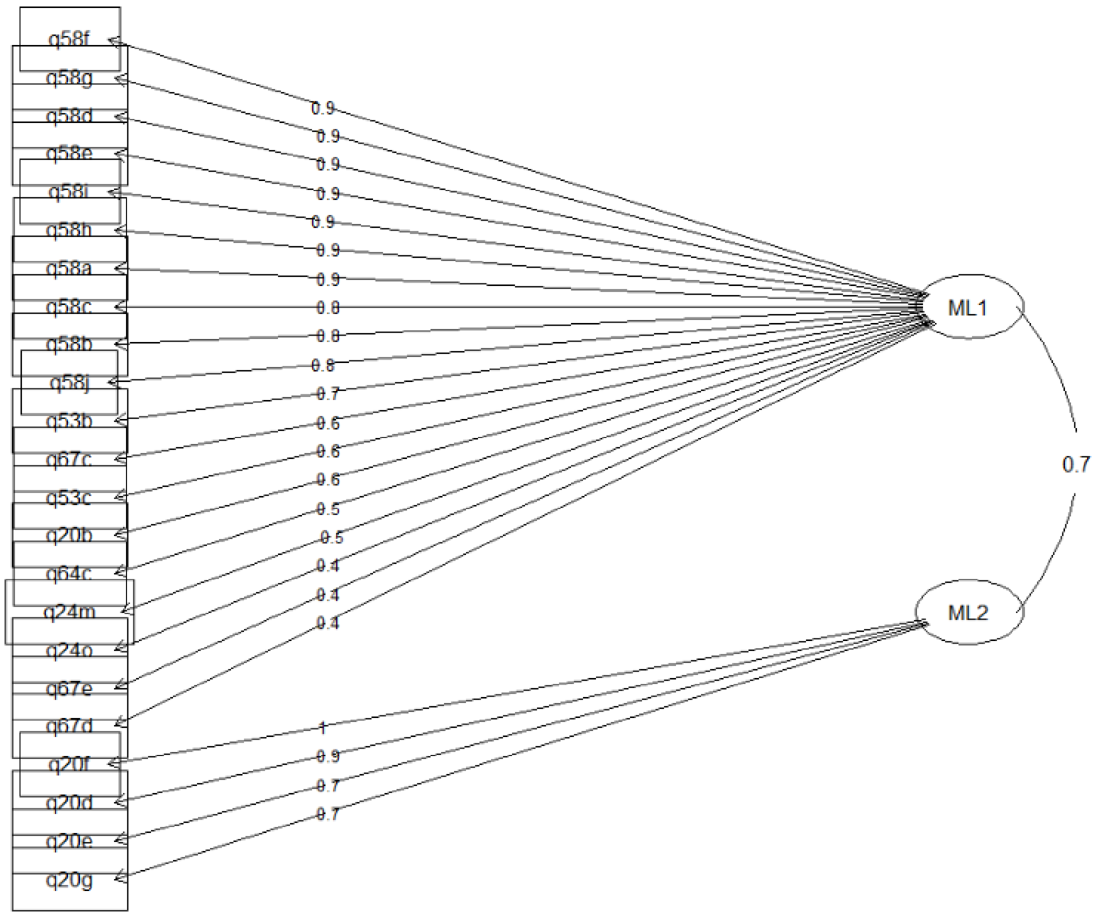


Observations:

- The first factor mainly focuses on Encouraging innovation, achieving results, self-learning etc. which are transformational leadership traits.
- Question in first factor that has low loading is Sick leaves as it is not a transformation trait.
- The second factor consists of traits of Consideration Leadership style
- The third factor seems to be invalid as it includes similar questions as that of factor 1 but has low loadings.
- The correlation between first two factor is quite high signifying overlap, but the third factor is very less correlated.

Based on above observation, we decided to develop the model with 2 factors to check the significance of 4 questions of ML3 above on the main two factors. The below model with same questions explains that:

- The questions of ML3 in above model are appearing with very low loadings. Hence, they should not be considered for our analysis.
- Although, we attempted to include these questions to find some other leadership styles, but due to unsatisfactory results we are dropping these variables for our next model.



Model 2: Supervisors Impressions

In this model, we have considered only those responses that reflect leadership of immediate supervisors. But before proceeding, we again ran a Data Validity & Assumption check to make sure the factor analysis will be valid.

```
$chisq
[1] 1821.068

$p.value
[1] 4.900764e-283

$df
[1] 153
```

Bartlett's Test

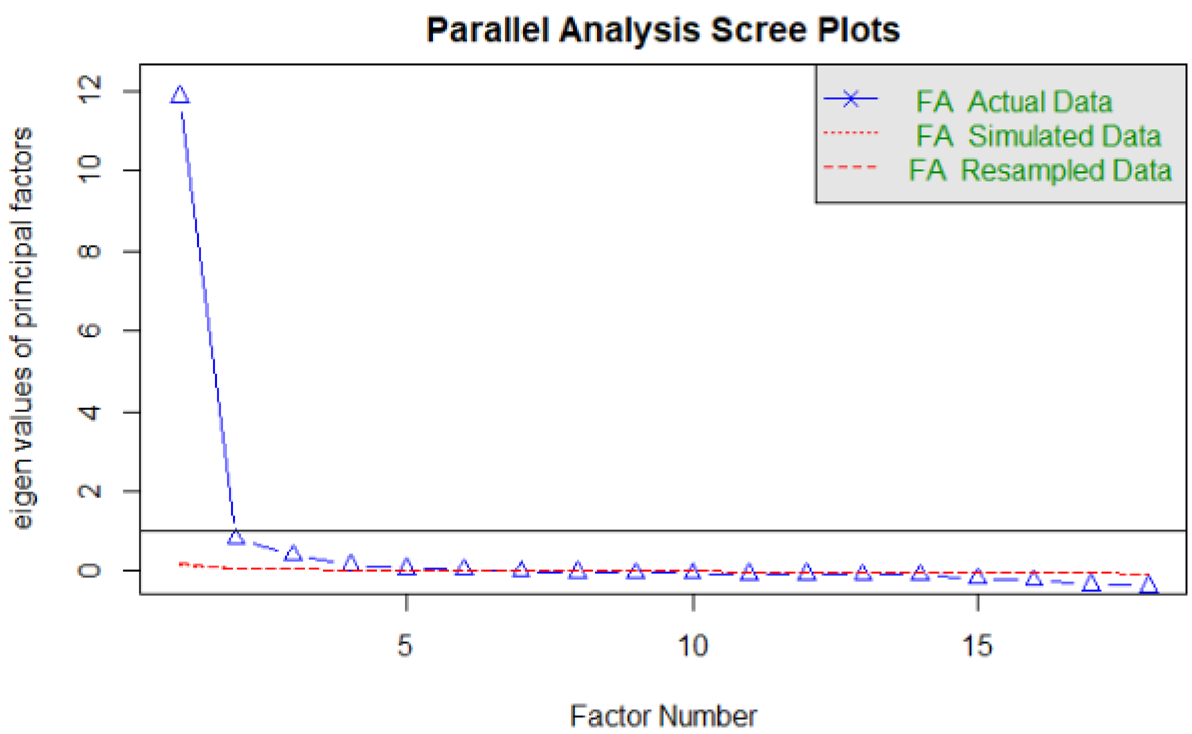
```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = correlation_red)
Overall MSA =  0.97
MSA for each item =
q20b q20d q20e q20f q20g q53b q53c q58a q58b q58c q58d q58e q58f q58g q58h q58i q58j
0.99 0.95 0.98 0.92 0.97 0.93 0.93 0.98 0.98 0.98 0.98 0.98 0.98 0.98 0.98 0.97 0.98
q67c
0.99
```

KMO Test

It can be observed that:

- Bartlett's Test results in p-value of 4.9 e-283 which is very less than 0.05. Thus, the variables have significant correlation and therefore are suitable for structure detection. It also helps us to select rotation for factor analysis model development. Since high correlation exists, we will select oblimin.
- KMO Test: The measure of Sampling Adequacy is 0.97 which is high and indicate high proportion of variance in variables that are caused by underlying factors. A factor analysis is thus, useful for such a dataset.

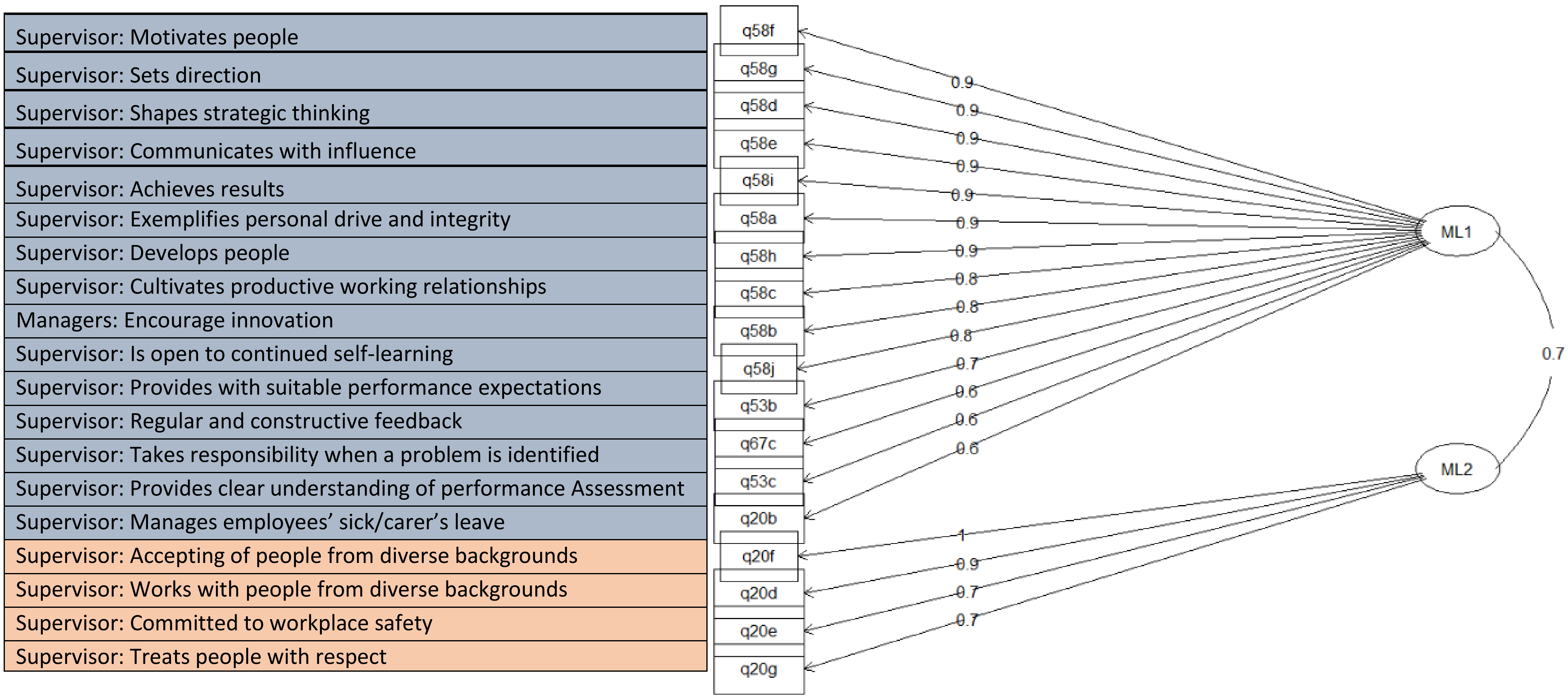
Below is the Parallel Analysis scree plot for this model describing eigen values of principal factors.



	Eigen_Values
1	11.869362576
2	0.809059949
3	0.408757665
4	0.160669905
5	0.062297583
6	0.051785068
7	-0.009339761
8	-0.029633006
9	-0.043234997
10	-0.044478369
11	-0.063648165
12	-0.073011012
13	-0.083383100
14	-0.090833320
15	-0.185812182
16	-0.226779965
17	-0.333934480
18	-0.374115041

It can be observed that based on New Kaiser Criterion, 2 factors have eigen values greater than 0.7. Thus, we are considering 2 factors for this model for conducting factor analysis.

Below is our final model for EFA.



Factor	Leadership Style	Characteristics of Leadership Style
ML1	**Transformational**	**Motivating & Setting Direction** **Encouraging innovation,** **Developing People & Shaping Strategic Thinking** **Achieving Results & Communicate with Influence** **Cultivating relationships,** **Open to self-learning**
ML2	**Consideration**	**Accept and work with diverse people,** **Commitment to workplace safety,** **Treating people with respect**

The model metrics are:

- Simple Structure : There is no double loading i.e., no survey question is being loaded on more than 1 factor. Thus, we have a simple Structure explaining the leadership styles.
- The cumulative variance being explained by the model is 65% which is good enough for considering these leadership styles as a way of predicting those employees that strive for creativity & innovation.
- The TLI for factoring reliability is quite high with a value of 0.913 indicating good fit.
- Although, RMSEA is 0.1 indicating mediocre fit, other metrics provide a fruitful result of this model.

The correlation between the two factors indicate a high overlap of traits of transformational & considerate leaders. This reflects that the Australian managers exhibit a mix of Transformational & Consideration Leadership skills.

```
Loadings:
      ML1    ML2
q20b  0.559
q20d         0.886
q20e         0.684
q20f         0.989
q20g         0.654
q53b  0.667
q53c  0.618
q58a  0.880
q58b  0.833
q58c  0.844
q58d  0.931
q58e  0.898
q58f  0.946
q58g  0.939
q58h  0.875
q58i  0.891
q58j  0.799
q64c  0.478
q67c  0.628

                 ML1   ML2
SS loadings    9.671 2.803
Proportion Var 0.509 0.148
Cumulative Var 0.509 0.656
```

```
Tucker Lewis Index of factoring reliability =  0.913
RMSEA index =  0.102  and the 90 % confidence intervals are  0.1 0.103
BIC =  16811.6
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                 ML1  ML2
Correlation of (regression) scores with factors   0.99 0.97
Multiple R square of scores with factors          0.98 0.94
Minimum correlation of possible factor scores     0.95 0.89
```

Below is the response distribution for selected question of EFA model.

It can be observed from responses that most of the respondents agree positively towards these questions. This proves the fact that Australia departments recruit supervisors who have transformation and consideration styles.

CORRESPONDANCE ANALYSIS

Correspondence analysis is a useful technique to visually understand the relationship between the categories of two categorical variables. It enables analyst to eliminate non-corresponding categories from a categorical variable thus reducing the number of rows/column in a grouped data. Although the dataset we are working on is ungrouped type, we have selected each question separately along with three grouping variables Agency Size, Agency Cluster & Age group to identify the correspondence between different categories of these groups and responses of employees.

Using these correspondence plots, we are trying to achieve the following objectives:

- Exploring differences between opinions of Male and Female Employees of their Immediate Workgroup & Supervisors
- Comparing perceptions of employees about Wellbeing & Productive/Ways of working in 2014 with those in 2015

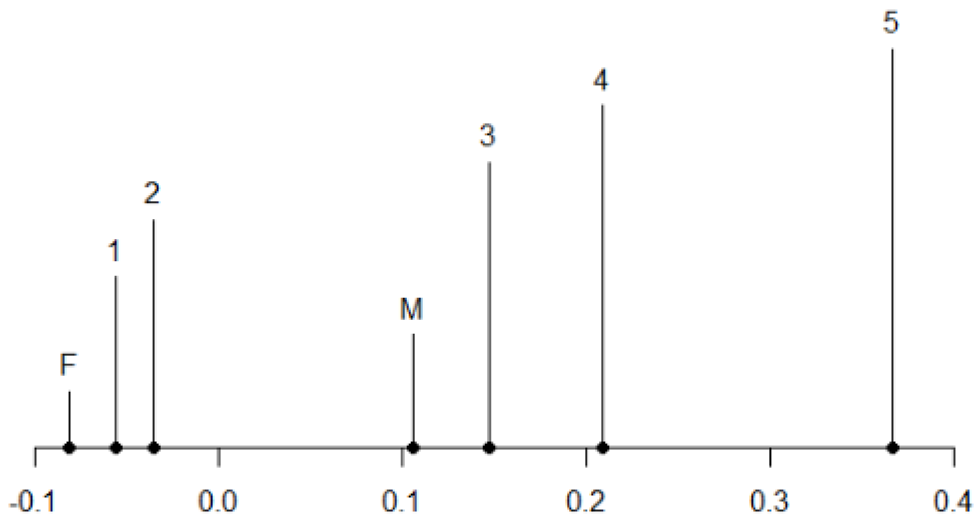
Since gender has only 2 categories Male and Female, the correspondence analysis results in a single dimension. Thus, a biplot is not feasible to analyse the correspondence. Thus, the following technique has been employed to visualize the CA:

- The row and column coordinates obtained from CA are stored in a single vector.
- The labels identifying these points as male, female and response levels are stored as a separate vector.
- The coordinates vector is plotted as a single number line with labels defined on each point to identify them

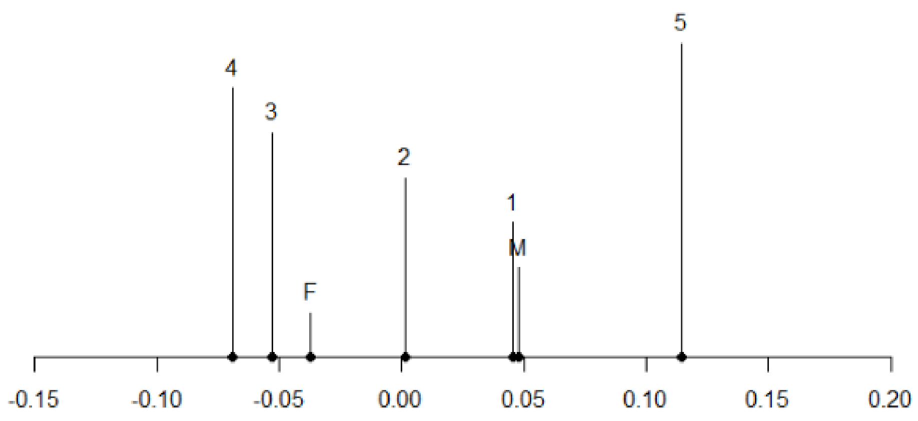
Through an iterative process, we explored all the combinations of Grouping and Response variables under the categories of Impressions of workgroup & immediate supervisors. Based on the exploration, we discovered two major correspondences for both categories of impressions.

Male Vs Female: Impressions of Immediate Workgroup

The plots describe the correspondence of male female with responses to following queries:



I have a clear understanding of how my workgroup's role contributes to my agency's strategic direction



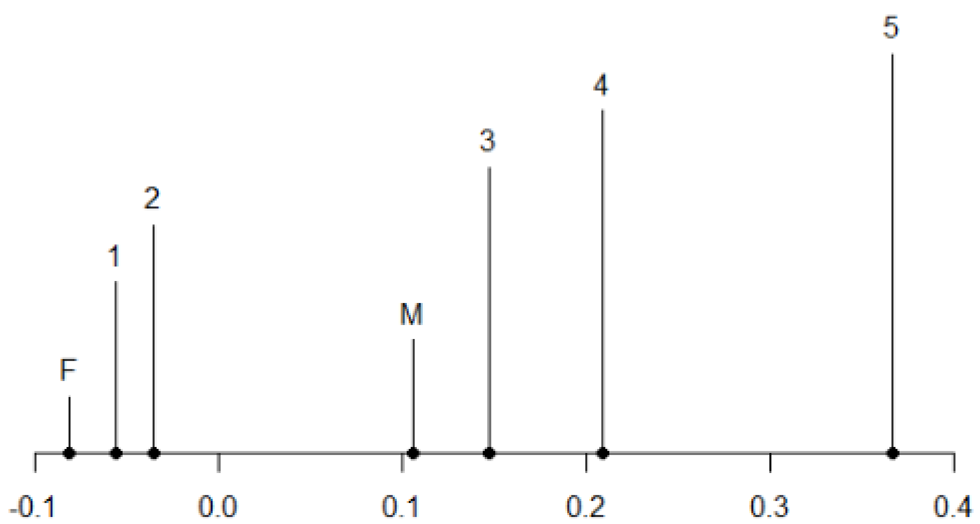
The people in my workgroup are honest, open and transparent in their dealings

It can be observed from these plots:

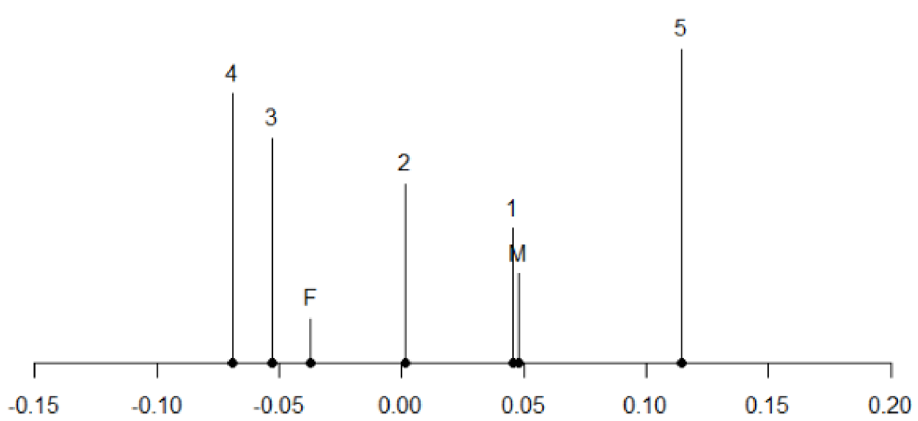
- Females have more clear understanding of workgroup's role contributes to agency's strategic direction than males
- Males believe that people in workgroup are honest, open and transparent in their dealings while females don't

Male Vs Female: Impressions of Immediate Supervisors

The plots describe the correspondence of male female with responses to following queries:



My supervisor works effectively with people from diverse backgrounds



My supervisor is committed to workplace safety

It can be observed from these plots:

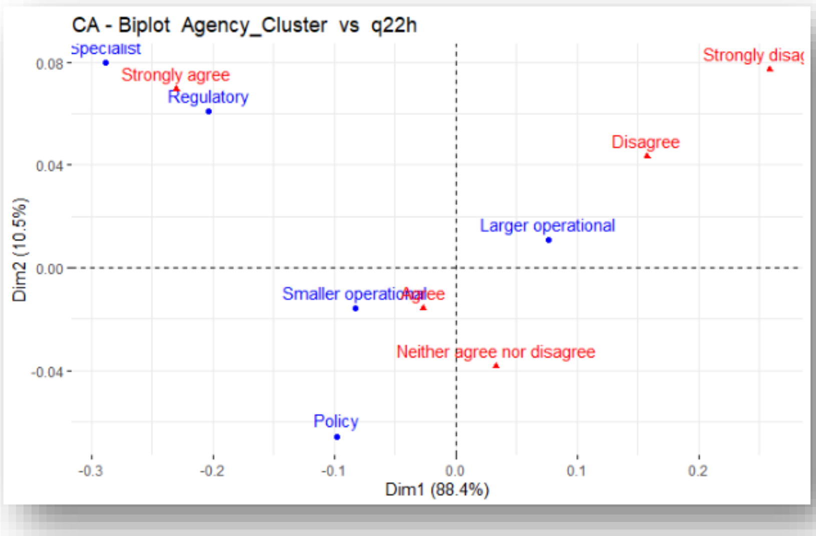
- Females witness supervisors working effectively with people from diverse backgrounds more than males
- Females believe that supervisors are committed to workplace safety while males don't

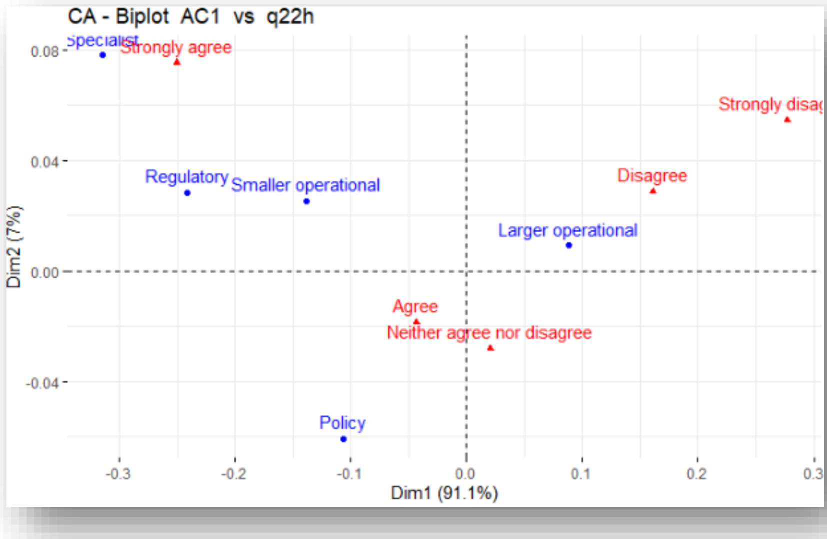
2014 Vs 2015: Perceptions of Employee regarding Wellbeing

The wellbeing of employees is an important factor for maintaining productivity, workplace effective environment and retention of employees. The Wellbeing section of APS survey consisted of various questions that reflect employee perceptions about wellbeing in their Agency. The following correspondence plots visualize some of the most important changes that can be seen from 2014 to 2015 regarding Wellbeing perceptions.

Agency genuinely cares about employees being healthy and safe at work

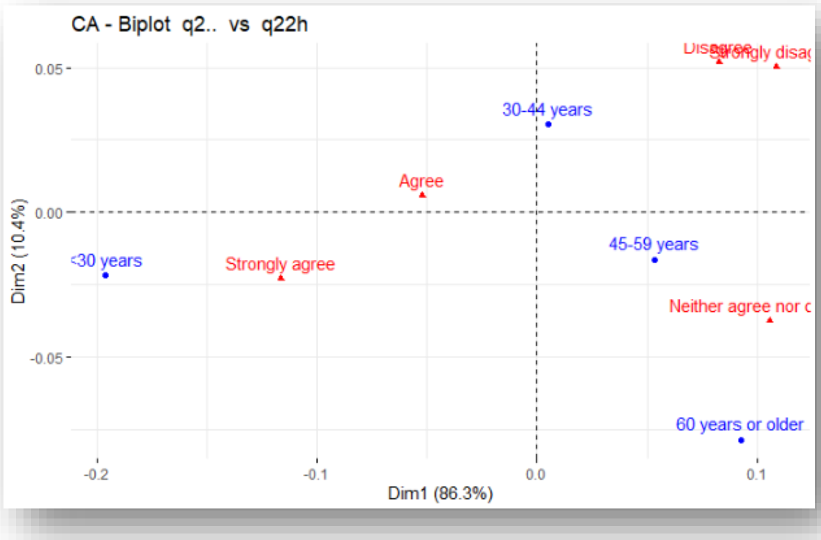
In various Agency Clusters:

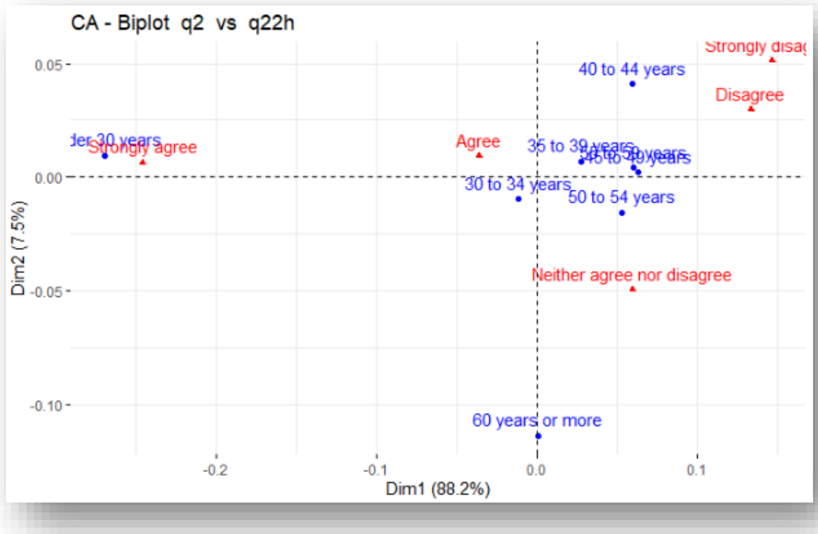




2014 **2015**

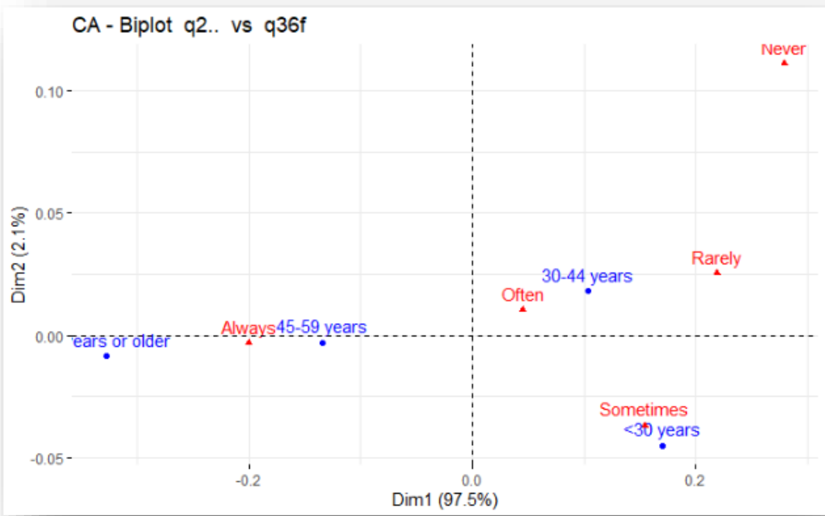
Among different Age Groups

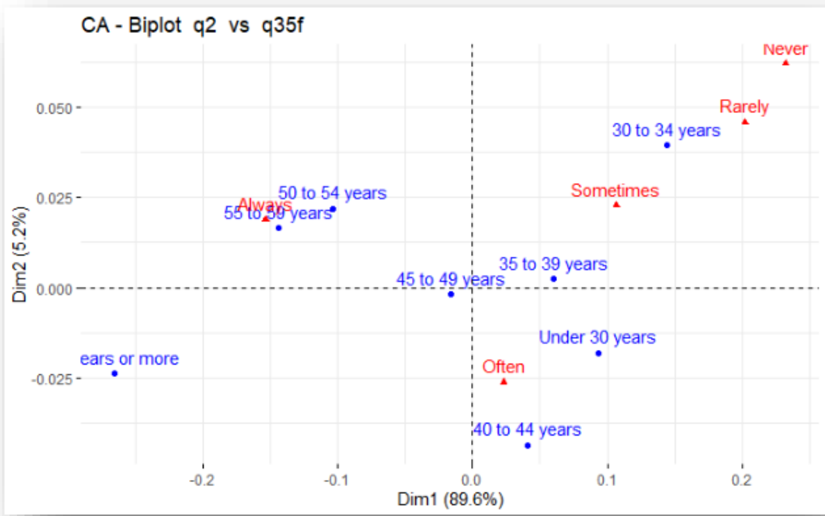




2014 **2015**

Clarity about duties & responsibilities within different Age Groups





2014 **2015**

Observations:

- The employee belief that agency cares about their health and safety decreased for Regulatory firms and Increased for Small operational firms from 2014 to 2015. Large operational received negative responses for both the years indicating no improvement in majority of agencies(since most of the agencies in dataset are large operational.

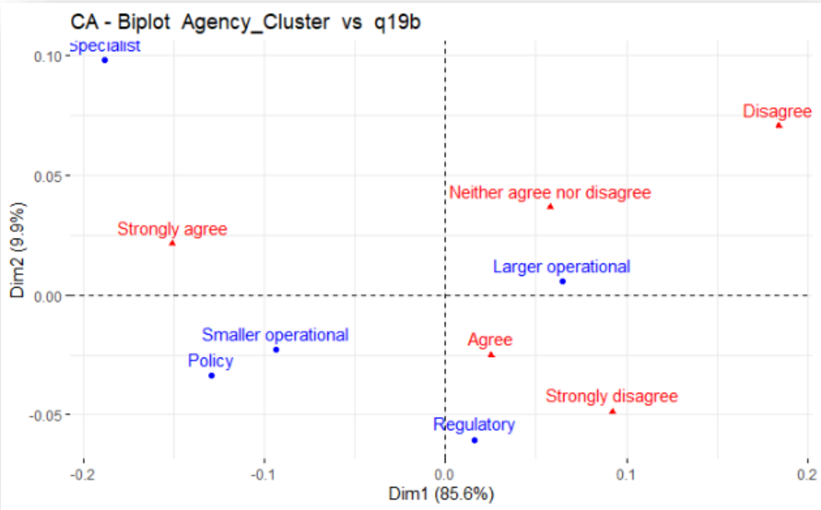
- The employees under 30 years strongly agreed to the notion that agency cared about their health and safety in 2015 while in 2014 the belief was not as strong.
- Employees of age 30-39 years were clearer about their duties and responsibilities in year 2014 as compared to 2015. However for people under 30 years this situation improved from 2014 to 2015.

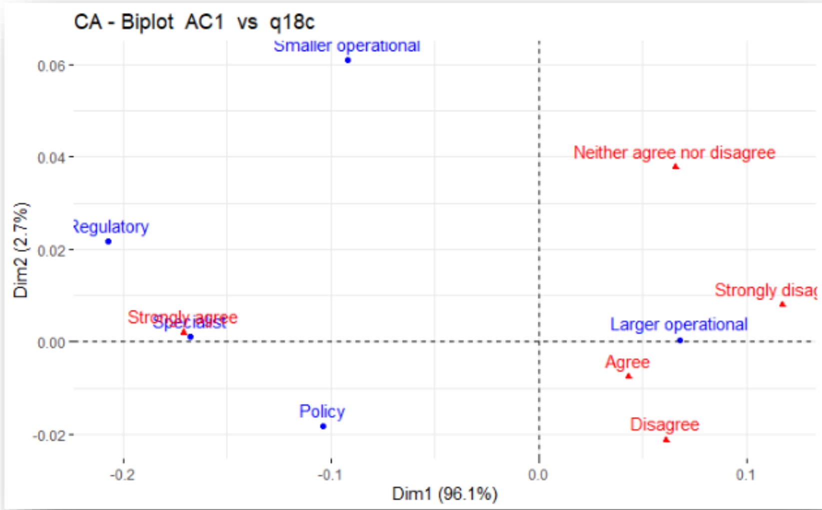
2014 Vs 2015: Perceptions of Employee regarding Productivity & Ways of Working

The productivity of employees ensures successful functioning of an organization. Managers always strive to come up with different ways of increasing productivity. It is directly related to methods, procedures, and measures in place for working effectively. The following correspondence plots visualize some of the most important changes that can be seen from 2014 to 2015 regarding perceptions of Productivity & ways of working.

Cooperation in workgroup to complete the jobs

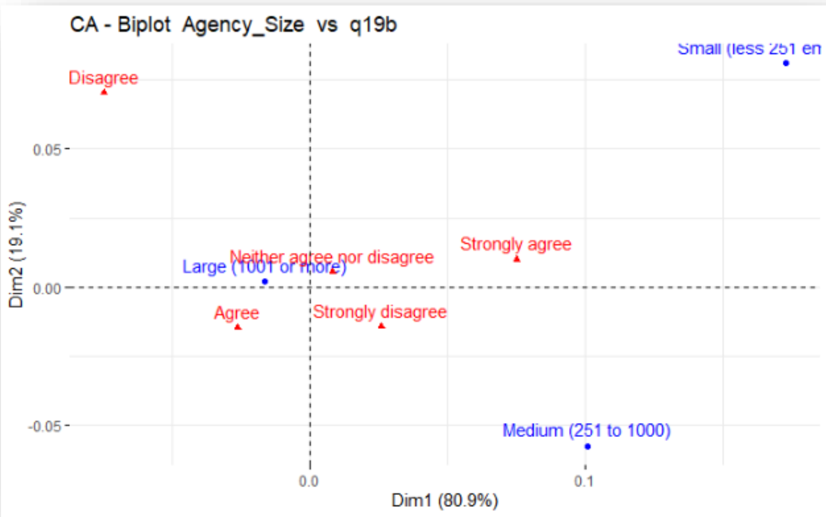
In different Agency Clusters

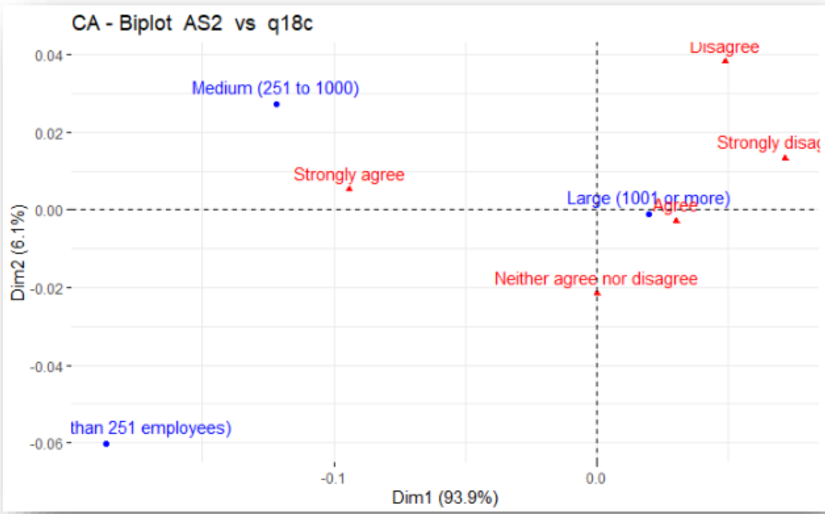




2014 **2015**

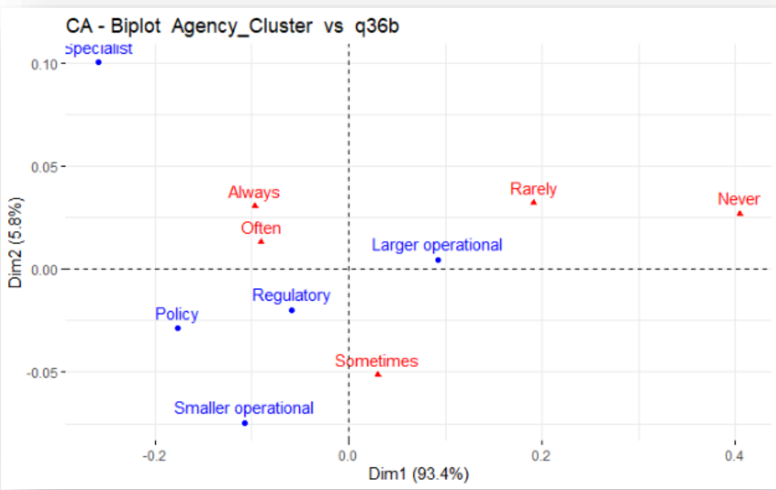
In different sized Agencies

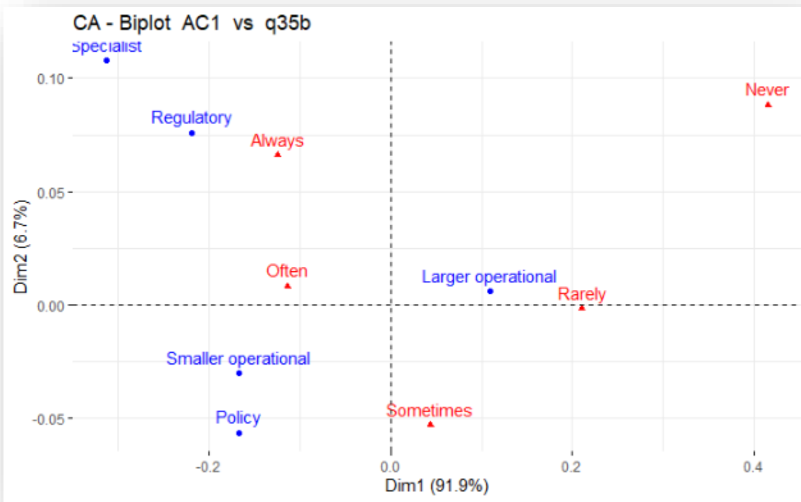




2014 **2015**

Choice of deciding ways of working in different agency clusters

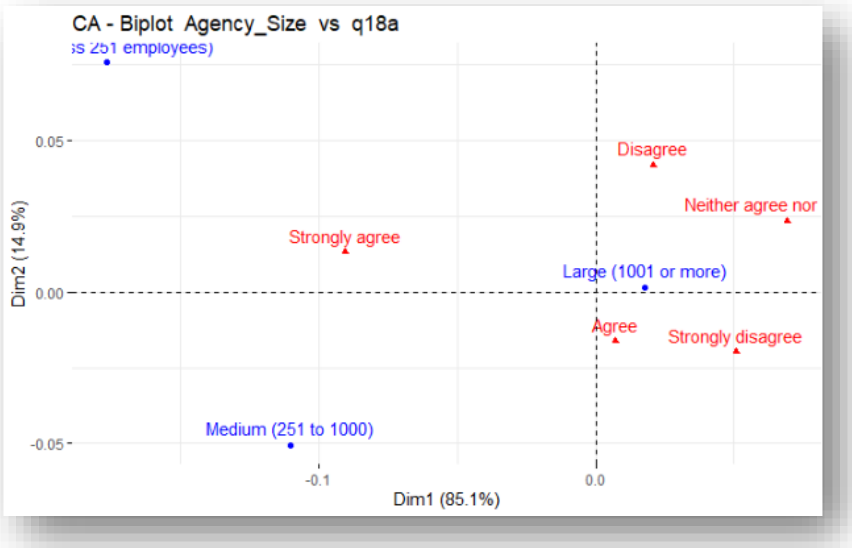


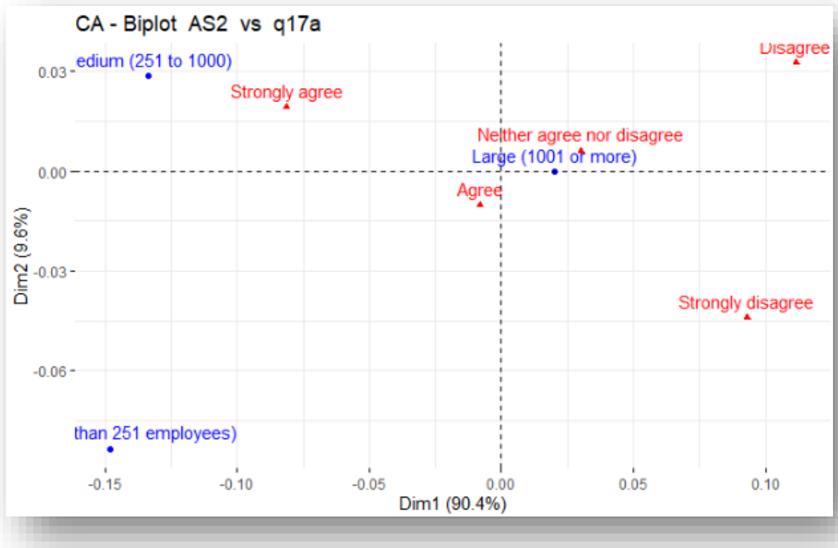


2014 **2015**

Enjoy working in current job

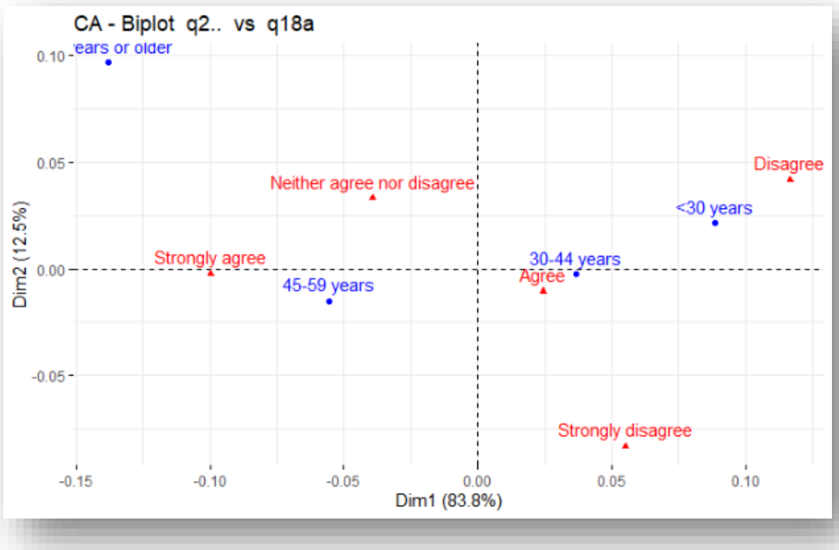
In different sized Agencies

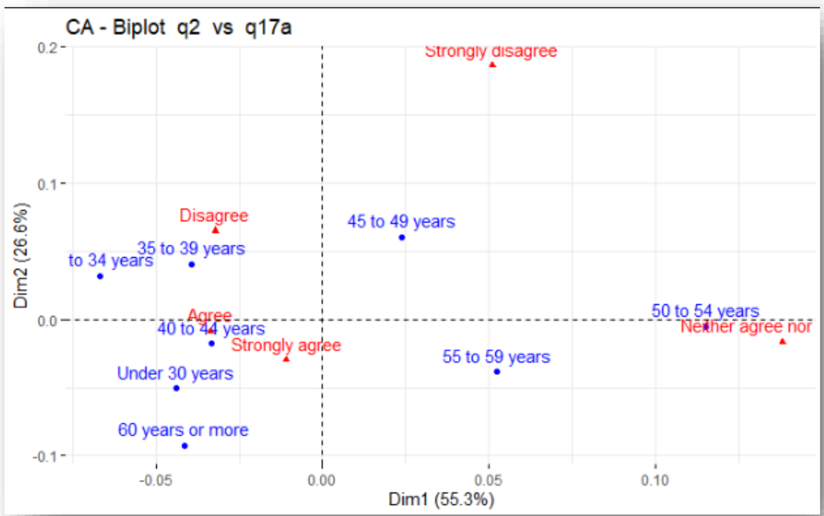




2014 **2015**

In different Age groups





2014 **2015**

Observations:

- Specialist Agencies received more positive response in terms of cooperation in workgroups for completing the job in 2015 as compared to 2014.
- Large Agencies also had a shift in response in terms of cooperation in workgroup from being neutral in 2014 to positive in 2015.
- Perceptions of getting a choice of deciding own way of working shifted from neutral to positive in Regulatory Agencies from 2014 to 2015.
- The medium sized agencies received highly positive feedback in 2015 as compared to 2014 regarding employees enjoying their current job.
- A dramatic shift in response from negative in 2014 to positive in year 2015 can be seen in employees of age less than 30 years regarding enjoying their current job.

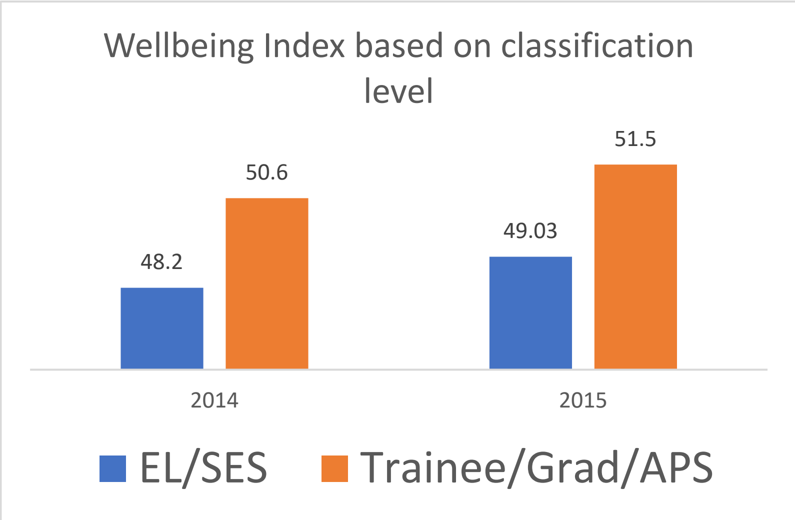
FURTHER INSIGHTS: WELLBEING & ENGAGEMENT INDEX

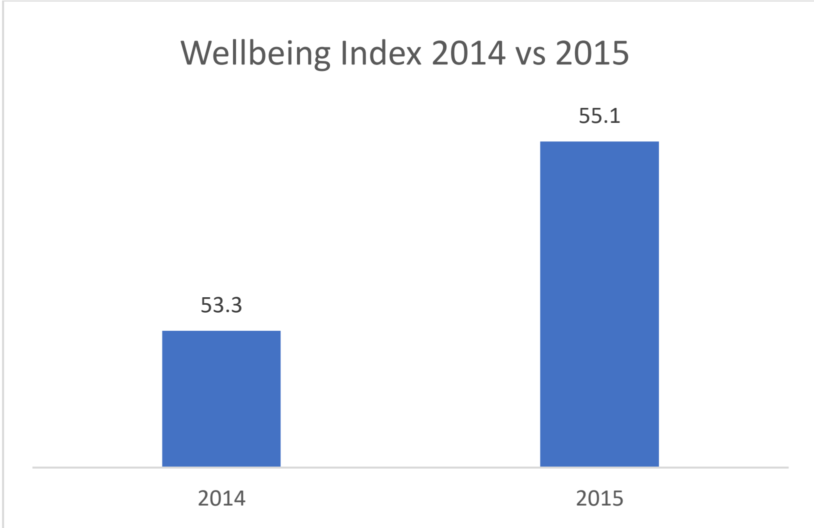
Employee Wellbeing Index

The wellbeing index provides a summarised view of employees towards their wellbeing in their organizations. The following queries are combined together based on responses to create Wellbeing index:

- My agency genuinely cares about employees being healthy and safe at work
- My agency supports employees who are injured or become ill due to work
- The wellbeing of our people is important
- Considering your work and life priorities, how satisfied are you with the work-life balance in your current job
- I have unrealistic time pressures
- I have a choice in deciding how I do my work
- I am clear what my duties and responsibilities are

The scale of the responses to these questions was changed to 10 as a regular practice of developing index. Then these were averaged across the selected questions for each respondent. The index is described in percentage and is averaged for all the employees to compare with the 2014 year.



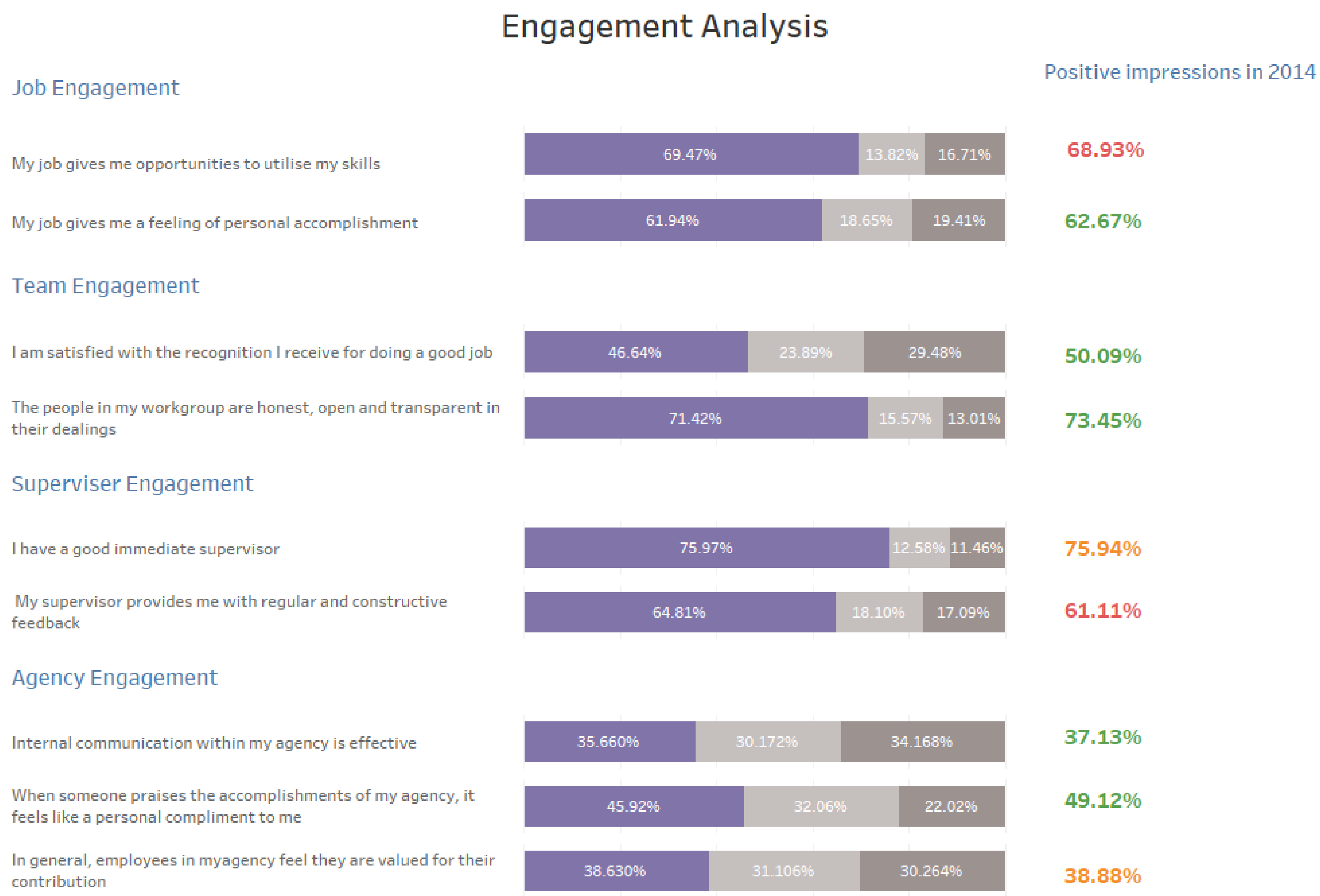


Observations:

- The wellbeing index slightly increased from 53% in year 2014 to 55% in year 2015 indicating a slight improvement in the situation of handling wellbeing in workplace in APS. Although it is not a drastic change, the change indicates a positive direction and scope for future development.
- Comparing based on classification levels, Trainees exhibit a better wellbeing index than Executive & Senior level employees. The fact that seniors have a lot of responsibilities to take care of the work pressure and timelines are very critical. Thus, we see a low wellbeing index for them.

Employee Engagement Index

Employee engagement refers to various metrics that evaluate managers and agencies interaction with employees. It is more than simply job satisfaction or commitment to an organisation. It is the extent to which employees are motivated, inspired and enabled to improve an organisation's outcomes. It is a two-way relationship that exists between an employee and their organisation. High levels of employee engagement is strongly associated with positive benefits such as increased performance and productivity.



The engagement analysis is divided into 4 categories. Each category has a set of questions that define the particular engagement. The below results were obtained from the analysis shown in above dashboard:

- The Job engagement more in terms of getting opportunity of utilizing skills and less in terms of feeling of accomplishment by employees in 2015 when compared to 2014.
- The Team Engagement decreased 2 to 4% in 2015 from 2014. Recognition received for a good job was less and people seemed less honest, open & transparent in their dealings in 2015
- Supervisor engagement enhanced slightly in 2015. Employee believed that immediate supervisors are good and provide regular constructive feedback.
- Agency Engagement decreased in 2015 from 2014 in terms of internal communications and personal attachment to company brand.

RECOMMENDATIONS

Based on our analysis i.e., Exploratory Factor analysis and Correspondence Analysis, we have defined the following recommendations for the Managers & Agencies of Australian Public Service that can help in increasing Productivity, Wellbeing, Retention and Overall Working culture.

- Supervisors should follow a gender equality in terms of their behavior towards employees.
- Larger Operational Agencies should develop more services and policies for taking care of wellbeing of the employees particularly for employee of more than 40 years of age
- Supervisors & agencies should develop well-structured plans and orientations for new hires and less experienced employees.
- Large Operational Agencies should device a structure that allows employees to clearly understand requirements and then have freedom to perform duties in their own ways.
- The Agencies & Supervisors should work on strategies of engaging with their employees in order to increase productivity.

CONCLUSION

This study provided an overall view of workplace culture, leadership, and productivity in Australian Public Service agencies in year 2015. We performed various analysis and techniques to gather insights from the APS survey dataset in an attempt to understand the scenario and provide useful recommendations for improvements in future. Below are the ultimate conclusions from this study:

- Majority of employees are working in Large Agencies with more than 1000 employees wherein Female population is more dominant with most employees belonging to 40 to 55 years.
- The data was found to be normal based on skewness & kurtosis metrics assuming it to be numeric data. The outliers identified in boxplot were not removed as they represent valid values.
- 23 questions were selected for EFA & CA that provided impressions of employees regarding leadership, productivity, wellbeing and work environment.
- The impressions regarding Regular & constructive feedback, encouraging innovation, clarity of performance expectations & owning responsibility of problem identified by seniors did not differ between supervisors & subordinates. However, their opinions differ significantly with managers being more positive about effectiveness in accepting & working with people from diverse background, commitment to workplace safety, receiving respect, getting informed of performance assessment, leadership skills, getting sick leaves and taking responsibilities of a problem, all by their supervisors and their agency valuing employees with encouragement to speak up on serious policy risks.
- The Exploratory Factor Analysis showed the presence of two major Leadership styles in Australian supervisors & managers. These are Transformational where the innovation, creativity, learning etc. is promoted, and Consideration where the wellbeing of employees is of concern. The model developed had a simple structure explaining 65% variation with a TLI of 0.913 and RMSEA of 0.1 indicating a mediocre fit.
- The correlation between the two factors indicated a high overlap of traits of transformational & considerate leaders. This reflects that the Australian managers exhibit a mix of Transformational & Consideration Leadership skills.
- The correspondence analysis is performed to explore differences between opinions of Male and Female Employees of their Immediate Workgroup & Supervisors and comparing perceptions of employees about Wellbeing & Productive/Ways of working in 2014 with those in 2015.
- Females witness supervisors working effectively with people from diverse backgrounds, committed to workplace safety more than males and have more clear understanding of workgroup's role contributes to agency's strategic direction than males. However, males believe that people in workgroup are honest, open and transparent in their dealings while females don't
- A better response was witnessed for specialist agencies for cooperation in workgroups for completing the job, for large agencies cooperation in workgroup, for medium sized agencies with people of age less than 30 being enjoyed by employees.

APPENDIX

PYTHON CODE:

```
#!/usr/bin/env python
# coding: utf-8

# # APS LEADERSHIP ANALYSIS

# ## Importing Libraries

from urllib.request import urlopen
import json
import numpy as np
import pandas as pd
import pandas as pd
import numpy as np
from scipy.stats import zscore
from scipy.stats import kurtosis, skew
import matplotlib.pyplot as plt
import scipy.stats as stats
pd.set_option('display.max_rows', 500)
pd.options.display.max_columns = 1000


# ## Getting Data from APS Data API

# ### Fetching 2015 Data : DO NOT RUN THIS CELL| WILL TAKE 20 MINS TO GET DATA

# APS 2015 API URL
url_2015 = 'https://data.gov.au/data/api/3/action/datastore_search?resource_id=0b7de79a-1355-47f1-add8-c2e0593dedbe'
fileobj_2015 = urlopen(url_2015)

# Loading JSON Data in Python
data_2015 = json.loads(fileobj_2015.read().decode())

# Creating Python Dataframe
df_2015 = pd.DataFrame(data_2015['result']['records'])

# Extracting Data from API as it gives only 100 records per iteration
while data_2015['result']['_links']['next'] is not None or data_2015['result']['_links']['next'] != '' or data_2015['result']['_links']['next'] != ' ':
    url_2015 = 'https://data.gov.au/data' + data_2015['result']['_links']['next']
    fileobj_2015 = urlopen(url_2015)
    data_2015 = json.loads(fileobj_2015.read().decode())
    df_2015 = pd.concat([df_2015, pd.DataFrame(data_2015['result']['records'])])
    if data_2015['result']['offset'] == 45200:
        break
```

```
# Removing Unwanted Fields
df_2015.drop(['_id', 'ResponseID', 'Portfolio', 'AS1', 'At_Classification'], axis=1, inplace=True)

# Saving Data to CSV so not to run API fetch every time
df_2015.to_csv('Sampled_Data.csv')




# ### Fetching 2014 Data  : DO NOT RUN THIS CELL| WILL TAKE 20 MINS TO GET DATA
# APS 2014 API URL
url_2014 = 'https://data.gov.au/data/api/3/action/datastore_search?resource_id=9b44e035-3bed-40dc-9687-34fc47b9f228'
fileobj_2014 = urlopen(url_2014)

# Loading JSON Data in Python
data_2014 = json.loads(fileobj_2014.read().decode())

# Creating Python Dataframe
df_2014 = pd.DataFrame(data_2014['result']['records'])

# Extracting Data from API as it gives only 100 records per iteration
while data_2014['result']['_links']['next'] is not None or data_2014['result']['_links']['next'] != '' or data_2014['result']['_links']['next'] != ' ':
    url_2014 = 'https://data.gov.au/data' + data_2014['result']['_links']['next']
    fileobj_2014 = urlopen(url_2014)
    data_2014 = json.loads(fileobj_2014.read().decode())
    df_2014 = pd.concat([df_2014, pd.DataFrame(data_2014['result']['records'])])
    if data_2014['result']['offset'] == 45200:
        break

# Saving Data to CSV so not to run API fetch every time
df_2015.to_csv('Sampled_Data_2014.csv')




# ## Handling Missing Data & Selecting Variables
aps_2015 = pd.read_csv('Sampled_Data.csv', na_values=[' '], low_memory=False)
columns_to_select_15 = ["AC1", "AS2","q1", "q2", "q7@", "q8","q17a","q17b","q17c","q17d","q18a",
                "q18b", "q18c", "q18d", "q18e", "q18f", "q19", "q20a", "q20b", "q20c",
                "q20d", "q20e", "q20f", "q20g", "q20h", "q20i", "q21a", "q21b", "q21c",
                "q21d", "q21e", "q21f", "q21g", "q21h", "q21i", "q21j","q22b","q22h","q22i",
                "q22j","q22k","q24c", "q24h","q24m", "q24o", "q24p", "q30", "q31", "q32",
                "q33", "q34a", "q34b", "q35a", "q35b", "q35c", "q35d", "q35e", "q35f", "q35g",
                "q36a", "q36b", "q37a", "q37b", "q37c", "q37d", "q37e", "q37f", "q37g", "q37h",
                "q37i", "q37j", "q37k", "q37l","q53b", "q53c", "q58a", "q58b", "q58c", "q58d",
                "q58e", "q58f", "q58g", "q58h", "q58i", "q58j", "q59", "q60", "q64c", "q67c",
                "q67d", "q67e", 'JobEngagement','TeamEngagement', 'SprvisrEngagement', 'AgencyEngagement']
```

```
aps_final_15 = aps_2015[columns_to_select_15]

## Extrating Missing Data
pd.DataFrame(aps_final_15.isna().sum()).to_csv('Missing_Data.csv')

## Filtering Missing data
aps_cleaned_2015 = aps_final_15[~aps_final_15.isnull().any(axis=1)]
aps_cleaned_2015.shape

# Saving Cleaned data
aps_cleaned_2015.to_csv('APS_FINAL_DATA.csv')

# Performing for 2014 dataset
aps_2014 = pd.read_csv('Sampled_Data_2014.csv', na_values=[' '], low_memory=False)
columns_to_select_2014 = ["Agency_Cluster", "Agency_Size", "q1", "q2@@", "Actual_Classification@",
                    "q8","q18a","q18b","q18c","q18d","q18e", "q19a", "q19b", "q19c", "q19d",
                    "q19e", "q20a", "q20b", "q20c", "q20d", "q20e", "q20f", "q20g", "q21a",
                    "q21b", "q21c", "q21d", "q21e", "q21f", "q21g", "q21h", "q21i", "q21j",
                    "q22b","q22h","q22i", "q22j", "q22o","q24c", "q24h","q24m", "q24o", "q24p",
                    "q31", "q32@", "q33", "q34", "q35a@", "q35b@", "q36a", "q36b", "q36c", "q36d",
                    "q36e", "q36f", "q36g", "q37a", "q37b", "q50b", "q50c", "q55a", "q55b", "q55c",
                    "q55d", "q55e", "q55f", "q55g", "q55h", "q55i", "q55j", "q56", "q57", "q79b",
                    "q79c", "q79d", "JobEngagement", "TeamEngagement", "SprvisrEngagement", "AgencyEngagementV2"]
aps_final_14 = aps_2014[columns_to_select_2014]
aps_cleaned_2014 = aps_final_14[~aps_final_14.isnull().any(axis=1)]
aps_cleaned_2014.shape
aps_cleaned_2014.to_csv('APS_FINAL_DATA_2014.csv')


# The endcoding of Responses from Strongly Agree, Agree etc. categories to numbers in correct order is performed in R.

# ## EFA Initialization

# ### Reading Encoded file  generated in R


# read encoded file
aps = pd.read_csv('APS_FINAL_DATA_2015_en.csv')
encode_df = pd.read_csv('APS_FINAL_DATA_2015_en.csv').iloc[:,2:]
encoded_df_14 = pd.read_csv('APS_FINAL_DATA_2014_en.csv').iloc[:,2:]


# ### Selecting Variables for EFA
# variable selection
efa_aps = aps[["q20b","q20d","q20e","q20f","q20g","q24m", "q24o","q53b", "q53c", "q58a",
          "q58b", "q58c", "q58d", "q58e", "q58f",
```

```
            "q58g", "q58h", "q58i", "q58j","q64c",
            "q67c","q67d", "q67e"]]


# ### Descriptive Analysis
# Descriptive Analysis

desc = efa_aps.agg([np.mean,np.std]).T
desc['skew'] = efa_aps.apply(skew)
desc['kurtosis'] = efa_aps.apply(kurtosis)
desc.to_csv("Descriptive Analysis.csv")

# ### Outlier Detection
# Outlier Detection

efa_aps_z =efa_aps.apply(zscore)
plt.figure(figsize=(10,8))
efa_aps_z.boxplot()
plt.figure(figsize=(10,8))
efa_aps.boxplot()

# ### Anova
# ANOVA
stat_df = pd.DataFrame({'Variable':[], 'F':[], 'p-value': [], 'Mean EL/SES': [],'Mean Trainee/Grad/APS':[], 'Mean Diff':[]} )
for col in ["q20b","q20d","q20e","q20f","q20g","q24m", "q24o","q53b", "q53c", "q58a",
          "q58b", "q58c", "q58d", "q58e", "q58f", "q58g", "q58h", "q58i", "q58j","q64c",
          "q67c","q67d", "q67e"]:

    owa = stats.f_oneway(encode_df[col][encode_df['q7.'] == 'EL/SES'],
            encode_df[col][encode_df['q7.'] == 'Trainee/Grad/APS'])

    mean1 = encode_df[col][encode_df['q7.'] == 'EL/SES'].mean()
    mean2 = encode_df[col][encode_df['q7.'] == 'Trainee/Grad/APS'].mean()

    stat_df = pd.concat([stat_df,pd.DataFrame([{'Variable':col, 'F':owa.statistic,
                                    'p-value': owa.pvalue, 'Mean EL/SES' : mean1,
                                    'Mean Trainee/Grad/APS': mean2, 'Mean Diff': abs(mean1-mean2)}])])
stat_df.to_csv('Anova.csv')


# ## Well Being Index: Change 2014 vs 2015
encode_df['wbi'] = (encode_df.loc[:,["q22h","q22i","q24c","q32","q35a","q35b","q35f"]] * 2).sum(axis=1)/7
encoded_df_14['wbi'] = (encoded_df_14.loc[:,["q22h","q22i","q24c","q33","q36a","q36b","q36f"]] * 2).sum(axis=1)/7


# ### Based on Agency Cluster
```

```
encode_df.groupby("AC1")["wbi"].mean()


# ### Based on Agency Size
encode_df.groupby("AS2")["wbi"].mean()


# ### Based on Agency Age Group
encode_df.groupby("q2")["wbi"].mean()


# ### Based on Experience
encode_df.groupby("q8")["wbi"].mean()


# ### Based on Workgroup Size
encode_df.groupby("q19")["wbi"].mean()


# ### Based on Supervisor Level
encode_df.groupby("q59")["wbi"].mean()


# ### Overall Mean
encode_df['wbi'].mean()
```

R Code:

```
# DANA 4830 Final Project

## Importing Libraries
```{r}
library(psych)
library(GPArotation)
library(FactoMineR)
library(factoextra)
library(ggplot2)
library(outliers)
library(tidyverse)
library(tidytext)
library(R.utils)
library(wordcloud)
library(viridis)
library(car)
library(dplyr)
```
```

Data Fetching, Cleaning and Variable Selection performed in Python

Reading Data Created in Python

```{r}

dataset_2014 <- read.csv("APS_FINAL_DATA_2014.csv")

dataset_2015 <- read.csv("APS_FINAL_DATA.csv")

```

Selecting Columns to Encode

```{r}

cols_2014 <- c("q18a","q18b", "q18c", "q18d", "q18e", "q19a", "q19b", "q19c", "q19d", "q19e", "q20a", "q20b", "q20c", "q20d", "q20e", "q20f", "q20g", "q21a", "q21b", "q21c", "q21d", "q21e", "q21f", "q21g", "q21h", "q21i", "q21j", "q22b", "q22h", "q22i", "q22j", "q22o", "q24h","q50b", "q50c", "q79b", "q79c", "q79d","q24c","q24m", "q24o", "q24p","q33", "q34", "q55a", "q55b", "q55c", "q55d", "q55e", "q55f", "q55g", "q55h", "q55i", "q55j","q36a", "q36b", "q36c", "q36d", "q36e", "q36f", "q36g")

cols_2015 <- c("q17a", "q17b", "q17c", "q17d", "q18a", "q18b", "q18c", "q18d", "q18e", "q18f", "q20a", "q20b", "q20c", "q20d", "q20e", "q20f", "q20g", "q20h", "q20i", "q21a", "q21b", "q21c", "q21d", "q21e", "q21f", "q21g", "q21h", "q21i", "q21j", "q22b", "q22h", "q22i", "q22j", "q22k", "q24h","q37a", "q37b", "q37c", "q37d", "q37e", "q37f", "q37g", "q37h", "q37i", "q37j", "q37k", "q37l", "q53b", "q53c", "q64c", "q67c", "q67d", "q67e","q24c", "q24m", "q24o", "q24p", "q32", "q33", "q58a", "q58b", "q58c", "q58d", "q58e", "q58f", "q58g", "q58h", "q58i", "q58j", "q35a", "q35b", "q35c", "q35d", "q35e", "q35f", "q35g")
```

Encoding Columns

```{r}

dataset_mod_2014 <- dataset_2014 %>%

        mutate_at(cols_2014,

            funs(recode(.,'"Strongly agree" = 1;"Agree" = 2;"Neither agree nor disagree" =3;"Disagree" = 4;"Strongly disagree" = 5')))

dataset_mod_2014 <- dataset_mod_2014 %>%

        mutate_at(cols_2014,

            funs(recode(.,'"To a very great extent" = 1;"Quite a lot" = 2;"Somewhat" =3;"Hardly at all" = 4;"Not at all" = 5')))

dataset_mod_2014 <- dataset_mod_2014 %>%

        mutate_at(cols_2014,

            funs(recode(.,'"Very satisfied" = 1;"Satisfied" = 2;"Neither satisfied nor dissatisfied" =3;"Dissatisfied" = 4;"Very dissatisfied" = 5')))

dataset_mod_2014 <- dataset_mod_2014 %>%

        mutate_at(cols_2014,

            funs(recode(.,'"Always" = 1;"Often"=2;"Sometimes"=3;"Rarely" = 4;"Never" = 5')))

dataset_mod_2015 <- dataset_2015 %>%

        mutate_at(cols_2015,

            funs(recode(.,'"Strongly agree" = 1;"Agree" = 2;"Neither agree nor disagree" =3;"Disagree" = 4;"Strongly disagree" = 5')))

dataset_mod_2015 <- dataset_mod_2015 %>%
```

~~~~
                mutate_at(cols_2015,
                        funs(recode(.,'"To a very great extent" = 1;"Quite a lot" = 2;"Somewhat" =3;"Hardly at all" = 4;"Not at all" = 5')))

dataset_mod_2015 <- dataset_mod_2015 %>%
                mutate_at(cols_2015,
                        funs(recode(.,'"Very satisfied" = 1;"Satisfied" = 2;"Neither satisfied or dissatisfied" =3;"Dissatisfied" = 4;"Very dissatisfied" = 5')))

dataset_mod_2015 <- dataset_mod_2015 %>%
                mutate_at(cols_2015,
                        funs(recode(.,'"Always" = 1;"Often"=2;"Sometimes"=3;"Rarely" = 4;"Never" = 5')))

```

# Negative Questions
```{r}
dataset_mod_2014 <- dataset_mod_2014 %>%
                mutate_at(c("q36a"),
                        funs(recode(.,'"Always" = 5;"Often"=4;"Sometimes"=3;"Rarely" = 2;"Never" = 1')))

dataset_mod_2015 <- dataset_mod_2015 %>%
                mutate_at(c("q35a"),
                        funs(recode(.,'"Always" = 5;"Often"=4;"Sometimes"=3;"Rarely" = 2;"Never" = 1')))

```

# Z Scaling
```{r}
dataZ = list()
for (i in c("q24m", "q24o","q53b", "q53c", "q58a", "q58b", "q58c", "q58d",
          "q58e", "q58f", "q58g", "q58h", "q58i", "q58j","q64c",
          "q67c","q67d", "q67e")){
  z.scores <- data.frame(dataset_mod_2015[i] %>% scores(type = "z"))
  dataZ[i] <- z.scores
}

max <-data.frame(mapply(max, dataZ))
min <-data.frame(mapply(min, dataZ))
max_minZ <- cbind(max,min)



z.scores <- dataset_mod_2015$q67e %>% scores(type = "z")
z.scores %>% summary()
length (which( abs(z.scores) >3 ))
```

# Writing to CSV Files
~~~~

```{r}
write.csv(max_minZ, "Zscore.csv")
write.csv(dataset_mod_2014, "APS_FINAL_DATA_2014_en.csv")
write.csv(dataset_mod_2015, "APS_FINAL_DATA_2015_en.csv")
```

Reading Encoded Data | Selecting Variables for EFA

```{r}
aps_data <- read.csv('APS_FINAL_DATA_2015_en.csv')

efa_data <- aps_data[c("q20b","q20d","q20e","q20f","q20g", "q24m", "q24o", "q53b", "q53c", "q58a", "q58b", "q58c", "q58d", "q58e","q58f", "q58g", "q58h", "q58i", "q58j","q64c", "q67c", "q67d", "q67e")]
aps_data$q2[aps_data$q2 ==  "45 to 49 years" | aps_data$q2 ==  "40 to 44 years" | aps_data$q2 ==  "50 to 54 years"]<- "40 to 54 years"
aps_data$q2[aps_data$q2 ==  "35 to 39 years" | aps_data$q2 ==  "30 to 34 years" | aps_data$q2 ==  "Under 30 years"]<- "Under 40 years"
aps_data$q2[aps_data$q2 != "40 to 54 years" & aps_data$q2 != "Under 40 years"] <- "55 years or more"

aps_data$q8[aps_data$q8 ==  "5 to less than 10 years" | aps_data$q8 ==  "1 to less than 5 years" | aps_data$q8 ==  "Less than 1 year"]<- "Less than 10 years"
aps_data$q8[aps_data$q8 ==  "10 to less than 15 years" | aps_data$q8 ==  "15 to less than 20 years" ]<- "10 to less than 20 years"
str(efa_data)
```

Exploratory Factor Analysis

Checking Correlation & Bartlett Test

```{r}
correlation <- cor(efa_data)
cortest.bartlett(correlation)
```

Since p-value is less than 0.05, thus variables are correlated thus we will use oblimin rotation for factor analysis

Checking KMO Test

```{r}
KMO(correlation)
```

KMO also results in correlation useful for ca for analysis

Determing Number of Factors to Select

```{r}
n_fact <- fa.parallel(efa_data, fm="ml", fa="fa")

```

One factor with large eigen value and other 5 with small values.

Model One: Supervisor + Agency

New Kaiser Criterion

```{r}
sum(n_fact$fa.values > 0.7)
```

```{r}
fa_model <- fa(efa_data, nfactors = 3, rotate = "oblimin", fm="ml")
fa.diagram(fa_model)
```

https://www.eaglesflight.com/blog/the-essential-role-of-feedback-in-leadership-development

```{r}
fa_model
```

Model Two: Supervisor only

```{r}
efa_data_red <- efa_data[-c(6,7,20,22,23)]
correlation_red <- cor(efa_data_red)
cortest.bartlett(correlation_red)
```

Since p-value is less than 0.05, thus variables are correlated thus we will use oblimin rotation for factor analysis

```{r}
KMO(correlation_red)
```

KMO also results in correlation useful for caftor analysis

```{r}
n_fact_red <- fa.parallel(efa_data_red, fm="ml", fa="fa")

```

One factor with large eigen value and other 5 with small values.

New Kaiser Criterion

```{r}
sum(n_fact_red$fa.values > 0.7)
```

```{r}
fa_model_red <- fa(efa_data_red, nfactors = 2, rotate = "oblimin", fm="ml")
fa.diagram(fa_model_red)
```

Simple Structure

```{r}
print(fa_model_red$loadings, cutoff = 0.3)
```

```{r}
fa_model_red
```

Mediocre FIT based on RMSEA

https://stats.idre.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis-confirmatory-factor-analysis/

Correspondance Analysis

Immediate Workgroup : Male vs Female

```{r}
for (col in c('q18a','q18b', 'q18c', 'q18d', 'q18e', 'q18f')){
  print(col)
  q.ca <- CA(table(aps_data$q1, aps_data[,col]))
  x_vals <- c(q.ca$row$coord,q.ca$col$coord)
  x_labs <- c('F','M','1','2','3','4','5')
  xlim <- c(min(x_vals)-0.1,max(x_vals)+0.1);
  ylim <- c(0,1);
  py <- c(0,0,0,0,0,0,0);
  ly <- c(0.1,0.2,0.3,0.4,0.5,0.6,0.7);

  ## create basic plot outline
  par(xaxs='i',yaxs='i',mar=c(5,1,1,1));
  plot(NA,xlim=xlim,ylim=ylim,axes=F,ann=F);
  axis(1);

  # ## plot elements
  segments(x_vals,py,x_vals,ly);
  points(x_vals,py,pch=16,xpd=NA);
  text(x_vals,ly,x_vals,pos=3, labels = x_labs);
  print(ggplot(data = as.data.frame(t(table(aps_data$q1, aps_data[,col]))), aes(x = Var2, y = Freq, fill = Var1)) +
  geom_bar(stat = "identity",
           position = position_dodge(),
           alpha = 0.75) +
  geom_text(aes(label = Freq),
            fontface = "bold",
            vjust = 1.5,
            position = position_dodge(.9),
            size = 4) +
  labs(x = "\n Gender", y = "Frequency\n",
       title = "\n Immediate Workgroup Analysis \n") +
```

  theme(plot.title = element_text(hjust = 0.5),
        axis.title.x = element_text(face="bold",
                                    colour="red",
                                    size = 12),
        axis.title.y = element_text(face="bold",
                                    colour="red",
                                    size = 12),
        legend.title = element_text(face="bold", size = 10)))
}

```

## Immediate Supervisor : Male vs Female

```{r}
library(ggplot2)
for (col in c('q20a','q20b', 'q20c', 'q20d', 'q20e', 'q20f', 'q20g', 'q20h','q20i')){
  print(col)
  q.ca <- CA(table(aps_data$q1, aps_data[,col]))
  x_vals <- c(q.ca$row$coord,q.ca$col$coord)
  x_labs <- c('F','M','1','2','3','4','5')
  xlim <- c(min(x_vals)-0.1,max(x_vals)+0.1);
  ylim <- c(0,1);
  py <- c(0,0,0,0,0,0,0);
  ly <- c(0.1,0.2,0.3,0.4,0.5,0.6,0.7);

  ## create basic plot outline
  par(xaxs='i',yaxs='i',mar=c(5,1,1,1));
  plot(NA,xlim=xlim,ylim=ylim,axes=F,ann=F);
  axis(1);

  # ## plot elements
  segments(x_vals,py,x_vals,ly);
  points(x_vals,py,pch=16,xpd=NA);
  text(x_vals,ly,x_vals,pos=3, labels = x_labs);
  print(ggplot(data = as.data.frame(t(table(aps_data$q1, aps_data[,col]))), aes(x = Var2, y = Freq, fill = Var1)) +
  geom_bar(stat = "identity",
           position = position_dodge(),
           alpha = 0.75) +
  geom_text(aes(label = Freq),
            fontface = "bold",
            vjust = 1.5,
            position = position_dodge(.9),
            size = 4) +
  labs(x = "\n Gender", y = "Frequency\n",
       title = "\n Immediate Supervisor Analysis \n") +

  theme(plot.title = element_text(hjust = 0.5),
        axis.title.x = element_text(face="bold",
                              colour="red",
                              size = 12),
        axis.title.y = element_text(face="bold",
                              colour="red",
                              size = 12),
        legend.title = element_text(face="bold", size = 10)))
}
```

Wellbeing Analysis : 2014 vs 2015

```{r}
aps_data_15 <- read.csv('APS_FINAL_DATA.csv')
for (col1 in c("AC1","AS2","q2")){
  for (col2 in c("q22h","q22i","q24c","q32","q35a","q35b","q35f"))
    {
    obs_res <- chisq.test(aps_data_15[,col1], aps_data_15[,col2])
    if (obs_res$p.value < 0.05 & col1 != col2){print(paste(col1,col2))
      print(paste(obs_res$statistic, obs_res$p.value))
      ca_model <- CA(table(aps_data_15[,col1], aps_data_15[,col2]), graph = FALSE)
      print(fviz_ca_biplot(ca_model, title = paste("CA - Biplot ", col1, " vs ", col2)))
      }
  }
}
```

Comparing with 2014
```{r}
aps_data_2014 <- read.csv('APS_FINAL_DATA_2014_en.csv')[,-c(1,2)]
```
```{r}
aps_data_14 <- read.csv('APS_FINAL_DATA_2014.csv')
for (col1 in c("Agency_Cluster", "Agency_Size", "q2..")){
  for (col2 in c("q22h","q22i","q24c","q33","q36a","q36b","q36f")){
    obs_res <- chisq.test(aps_data_14[,col1], aps_data_14[,col2])
    if (obs_res$p.value < 0.05 & col1 != col2){print(paste(col1,col2))
      print(paste(obs_res$statistic, obs_res$p.value))
      ca_model <- CA(table(aps_data_14[,col1], aps_data_14[,col2]), graph = FALSE)
      print(fviz_ca_biplot(ca_model, title = paste("CA - Biplot ", col1, " vs ", col2)))
      }
  }
}

````
```

## Productivity & Ways of Working Analysis : 214 vs 2015

```{r}
for (col1 in c("AC1","AS2","q2")){
  for (col2 in c("q17a","q18c","q24h","q35b","q35f"))
    {
    obs_res <- chisq.test(aps_data_15[,col1], aps_data_15[,col2])
    if (obs_res$p.value < 0.05 & col1 != col2){print(paste(col1,col2))
      print(paste(obs_res$statistic, obs_res$p.value))
      ca_model <- CA(table(aps_data_15[,col1], aps_data_15[,col2]), graph = FALSE)
      print(fviz_ca_biplot(ca_model, title = paste("CA - Biplot ", col1, " vs ", col2)))
      }
  }
}

```

Comparing with 2014
```{r}
aps_data_14 <- read.csv('APS_FINAL_DATA_2014.csv')
```
```{r}
for (col1 in c("Agency_Cluster", "Agency_Size", "q2..")){
  for (col2 in c("q18a","q19b","q24h","q36b","q36f")){
    obs_res <- chisq.test(aps_data_14[,col1], aps_data_14[,col2])
    if (obs_res$p.value < 0.05 & col1 != col2){print(paste(col1,col2))
      print(paste(obs_res$statistic, obs_res$p.value))
      ca_model <- CA(table(aps_data_14[,col1], aps_data_14[,col2]), graph = FALSE)
      print(fviz_ca_biplot(ca_model, title = paste("CA - Biplot ", col1, " vs ", col2)))
      }
  }
}

```

Further Analysis in Python Based on Index
````
````