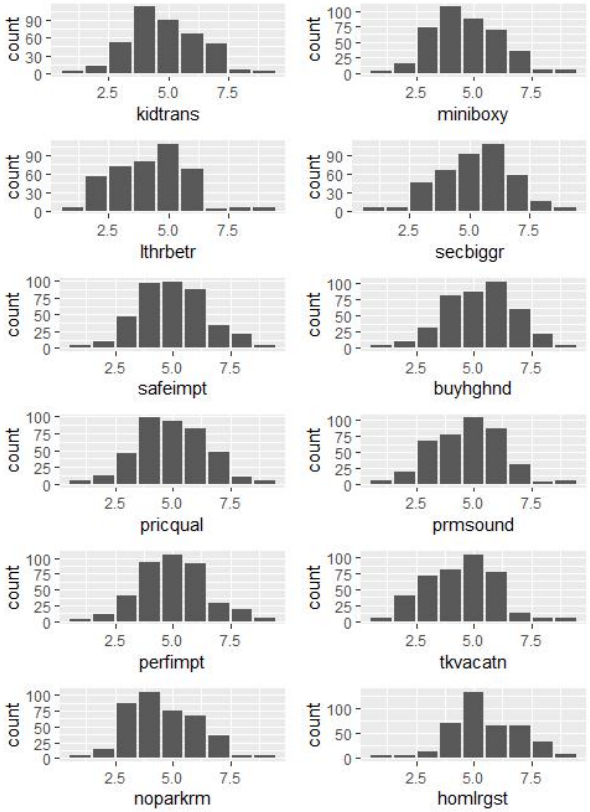
Microvan Case

Yiwen(Julie) Zhang

Exploratory Data Analysis (EDA)

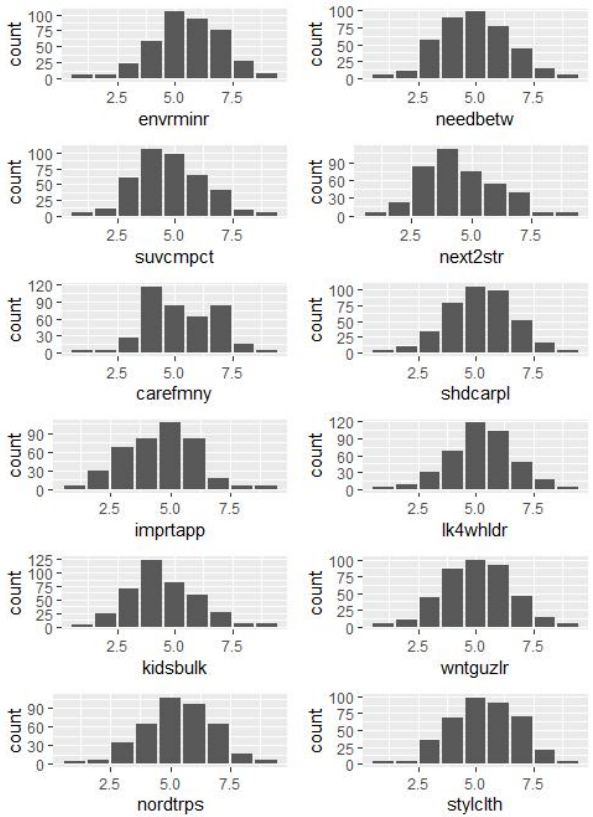


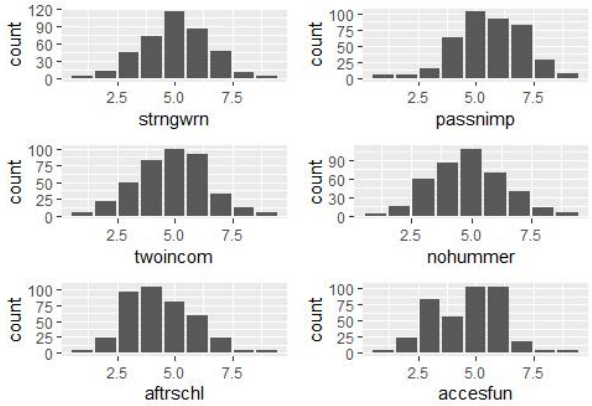
Most of the distributions of the 30 variables resembled to normal distribution closely:

- They have means around 5 (the center)
- The curves are roughly symmetric at the center

Exceptions being the variable “carefmny” and “accesfun” of which distributions are less symmetric. But that was not much of a concern since majority of their values are within 2 to 6.

EDA





Neither did we find outliers by examining the histogram of variables.

We also took a closer look at the overall row mean and row sum. We did not find any respondent who showed no understanding of the survey.

Full Model

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.381266   2.958551   0.129   0.8975
kidtrans     0.241114   0.164793   1.463   0.1443
miniboxy     0.177881   0.129258   1.376   0.1696
lthrbetr     0.247630   0.121971   2.030   0.0430 *
secbiggr    -0.104796   0.105833  -0.990   0.3227
safeimpt    -0.018525   0.133630  -0.139   0.8898
buyhghnd     0.112578   0.116162   0.969   0.3331
pricqual     0.105322   0.104796   1.005   0.3155
prmsound     0.010118   0.108312   0.093   0.9256
perfimpt     0.232663   0.128198   1.815   0.0704 .
tkvacatn     0.166171   0.124671   1.333   0.1834
noparkrm     0.178143   0.115804   1.538   0.1248
homlrgst    -0.208684   0.122418  -1.705   0.0891 .
envrminr    -0.033245   0.122777  -0.271   0.7867
needbetw     0.128468   0.102636   1.252   0.2115
suvcmpct     0.215136   0.122643   1.754   0.0802 .
next2str     0.024294   0.106843   0.227   0.8203
carefmny    -0.243143   0.134373  -1.809   0.0712 .
shdcarpl    -0.286783   0.122413  -2.343   0.0197 *
imprtapp     0.059086   0.104214   0.567   0.5711
lk4whldr    -0.064119   0.126739  -0.506   0.6132
kidsbulk    -0.096959   0.122063  -0.794   0.4275
wntguzlr    -0.028943   0.115689  -0.250   0.8026
nordtrps     0.073056   0.127473   0.573   0.5669
stylclth     0.015757   0.113597   0.139   0.8898
strngwrn    -0.196806   0.113448  -1.735   0.0836 .
passnimp     0.161975   0.119056   1.360   0.1745
twoincom     0.170419   0.096469   1.767   0.0781 .
nohummer     0.009052   0.095697   0.095   0.9247
aftrschl    -0.025716   0.116551  -0.221   0.8255
accesfun    -0.003458   0.122112  -0.028   0.9774
```

- All 400 records of 30 variables are in this full regression model.
- In general, all signs on coefficients are making sense to a concept of microvan. In particular,
  - ***Signs that support concept of microvan in car size*** - nordtrps / suvcmpct/ needbetw / noparkrm / homlrgst
  - ***Signs of personality of the supporters of microvan*** **-** twoincom / imprtapp / stylclth / passnimp / carefmny
  - ***Signs that indicates uniqueness of microvan features*** - lthrbetr / buyhghnd / pricequal / prmsound
  - Only few exceptions that seems contradicting with the concept of microvan
    - secbiggr (negative) - likes wants bigger car
    - perfimpt (positive) - actually concerns power/ acceleration
    - kidsbulk (negative) - need bigger car to hold kid's stuff
    - aftrschl (negative) - kids generally have more after-school activities
- However, **the important fact is**, only too few (2) of these coefficients are statistically significant by p values. We will hope to conduct factor analysis to identify and take away any correlations between some of these variables and narrow down to a few factors for this study.

Factor Analysis - Evaluate Data

```
        Bartlett's Test of Sphericity

Call: bart_spher(x = data[, 3:32])

     X2 = 7884.075
     df = 435
p-value < 2.22e-16
```

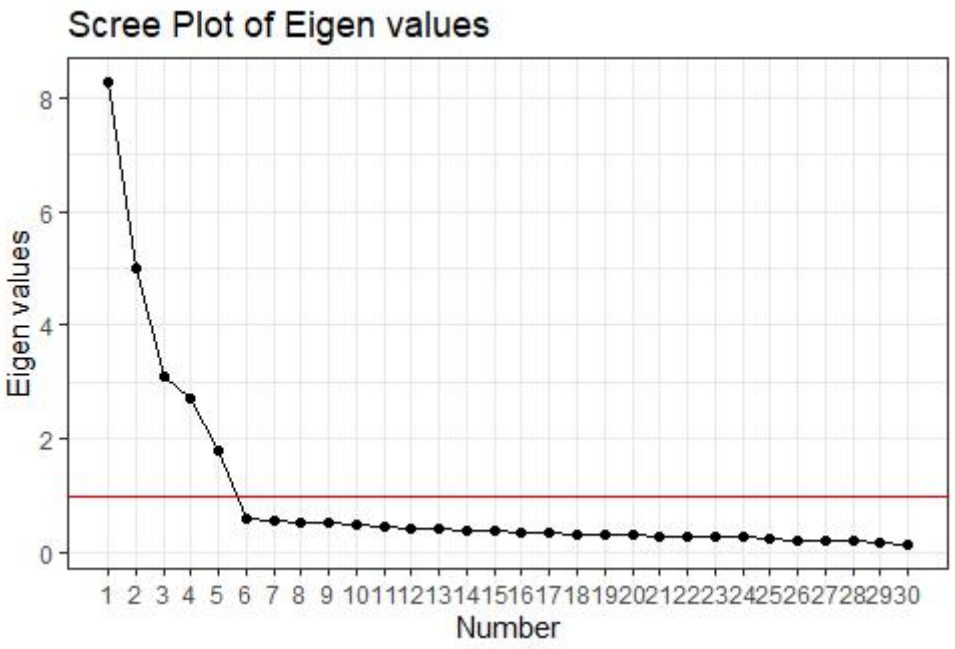
```
Kaiser-Meyer-Olkin Statistics

Call: KMOS(x = data[, 3:32])

Measures of Sampling Adequacy (MSA):
 kidtrans  miniboxy  lthrbetr  secbiggr  safeimpt  buyhghnd  pricqual  prmsound  perfimpt
0.8061785 0.9160429 0.9678512 0.9202145 0.7779786 0.9427124 0.9116132 0.9660718 0.7785772
 tkvacatn  noparkrm  homlrgst  envrminr  needbetw  suvcmpct  next2str  carefmny  shdcarpl
0.9689191 0.9269014 0.9414695 0.8545773 0.9309465 0.9272547 0.9209833 0.9569440 0.8609554
 imprtapp  lk4whldr  kidsbulk  wntguzlr  nordtrps  stylclth  strngwrn  passnimp  twoincom
0.9693660 0.8576882 0.9003552 0.9326443 0.8800392 0.9627968 0.9012162 0.9647283 0.9402597
 nohummer  aftrschl  accesfun
0.9264645 0.9052630 0.9690896

KMO-Criterion: 0.9233418
```

- Bartlett's Test of Sphericity: p-value<0.05
- At least some of the variables have strong correlations, and thus correlation matrix is different from identity matrix.

- KMO values are all above 0.6
- The proportion of variance being common variance is low. Samples are adequate.

- Conclusion: we can proceed to factor analysis.

Factor Analysis - Eigen Values



- 5 factors have Eigen values > 1, so they convey more information than an individual variable.
- We tried 6 factors as cutoffs. It turned out that the 6th factor has Eigen value of 0.63 and it’s only highly correlated with variable “nohummer”.
- Therefore, we chose 5 as number of factors.

Factor Analysis - RC1

	RC1	RC2	RC3	RC5	RC4
kidtrans	0.1237	0.0023	0.9330	-0.0196	0.0067
safeimpt	0.0261	0.0465	0.0536	-0.0197	0.9075
perfimpt	0.1111	-0.0810	-0.0756	0.0256	-0.8836
envrminr	-0.1734	-0.0349	0.0855	-0.8666	-0.0136
shdcarpl	0.1636	-0.0277	-0.0551	0.8665	0.0775
lk4whldr	0.1676	0.0247	0.0318	0.1041	0.8556
nordtrps	-0.0637	-0.1018	-0.8665	-0.0135	-0.0410
carefmny	-0.7634	-0.1452	-0.1975	-0.3081	-0.0805
miniboxy	0.1229	0.8419	-0.1072	0.0511	0.0131
suvcmpct	0.0838	0.8191	0.2022	0.0371	-0.0036
kidsbulk	0.1799	0.0173	0.8248	0.0573	0.0225
wntguzlr	-0.3571	0.0323	-0.0074	-0.7628	0.0150
buyhghnd	0.8147	0.1758	0.0222	0.0541	0.0981
tkvacatn	0.6539	-0.0310	0.2570	0.4583	0.0225
homlrgst	0.3309	-0.6794	0.1535	0.3166	0.0855
aftrschl	0.1951	-0.1092	0.7754	-0.1132	0.1819
strngwrn	0.2730	-0.2587	0.0828	0.0609	0.7354
lthrbetr	0.7107	-0.1865	0.2461	0.2897	0.0675
noparkrm	0.1739	0.8066	0.0121	-0.0872	-0.0164
accesfun	0.6784	-0.0387	0.3008	0.3695	0.0025
pricqual	0.7823	-0.1904	-0.0821	-0.1353	0.0041
passnimp	-0.6475	-0.0197	-0.3995	-0.2767	0.0143
next2str	0.2581	-0.7429	0.1056	-0.1152	-0.0068
stylclth	0.6036	0.2363	0.1824	0.4262	-0.0270
twoincom	0.7562	0.1170	-0.0943	-0.0741	0.0951
secbiggr	-0.0759	0.7593	0.0587	0.0332	-0.0825
needbetw	0.1264	0.7575	-0.0111	0.0434	0.0424
prmsound	0.6819	-0.0175	0.1678	0.2897	0.0731
imprtapp	0.5094	-0.0102	0.3456	0.3495	0.2000
nohummer	0.0595	0.7064	0.0455	-0.0376	0.0384

- Factor 1 has high **negative correlation** with carefmny (Careful with money) and high **positive correlation** with buyghgnd (buy higher-end cars), pricqual (Car prices strongly reflect underlying production quality), lthrbetr (Leather seats are dramatically better than cloth), and twoincom (Our family would find it hard to subsist on just one income).

- Factor 1 is related to **price premium for quality.**

Factor Analysis - RC2

	RC1	RC2	RC3	RC5	RC4
kidtrans	0.1237	0.0023	0.9330	-0.0196	0.0067
safeimpt	0.0261	0.0465	0.0536	-0.0197	0.9075
perfimpt	0.1111	-0.0810	-0.0756	0.0256	-0.8836
envrminr	-0.1734	-0.0349	0.0855	-0.8666	-0.0136
shdcarpl	0.1636	-0.0277	-0.0551	0.8665	0.0775
lk4whldr	0.1676	0.0247	0.0318	0.1041	0.8556
nordtrps	-0.0637	-0.1018	-0.8665	-0.0135	-0.0410
carefmny	-0.7634	-0.1452	-0.1975	-0.3081	-0.0805
miniboxy	0.1229	0.8419	-0.1072	0.0511	0.0131
suvcmpct	0.0838	0.8191	0.2022	0.0371	-0.0036
kidsbulk	0.1799	0.0173	0.8248	0.0573	0.0225
wntguzlr	-0.3571	0.0323	-0.0074	-0.7628	0.0150
buyhghnd	0.8147	0.1758	0.0222	0.0541	0.0981
tkvacatn	0.6539	-0.0310	0.2570	0.4583	0.0225
homlrgst	0.3309	-0.6794	0.1535	0.3166	0.0855
aftrschl	0.1951	-0.1092	0.7754	-0.1132	0.1819
strngwrn	0.2730	-0.2587	0.0828	0.0609	0.7354
lthrbetr	0.7107	-0.1865	0.2461	0.2897	0.0675
noparkrm	0.1739	0.8066	0.0121	-0.0872	-0.0164
accesfun	0.6784	-0.0387	0.3008	0.3695	0.0025
pricqual	0.7823	-0.1904	-0.0821	-0.1353	0.0041
passnimp	-0.6475	-0.0197	-0.3995	-0.2767	0.0143
next2str	0.2581	-0.7429	0.1056	-0.1152	-0.0068
stylclth	0.6036	0.2363	0.1824	0.4262	-0.0270
twoincom	0.7562	0.1170	-0.0943	-0.0741	0.0951
secbiggr	-0.0759	0.7593	0.0587	0.0332	-0.0825
needbetw	0.1264	0.7575	-0.0111	0.0434	0.0424
prmsound	0.6819	-0.0175	0.1678	0.2897	0.0731
imprtapp	0.5094	-0.0102	0.3456	0.3495	0.2000
nohummer	0.0595	0.7064	0.0455	-0.0376	0.0384

- Factor 2 has high **negative correlation** with next2str (Next car will be a two-seater) and high **positive correlation** with miniboxy (Current minivans are too boxy and large), suvcmpct (Like SUVs more than minivans since they're more compact), noparkrm (Current residence doesn't have a lot of parking room), secbiggr (Second car would need to be bigger than a standard sedan), needbetw (Needs to be something between a sedan and a minivan) and nohummer (Not interested in owning a vehicle like a Hummer)

- Factor 2 is related to **medium car size**

Factor Analysis - RC3

	RC1	RC2	RC3	RC5	RC4
kidtrans	0.1237	0.0023	0.9330	-0.0196	0.0067
safeimpt	0.0261	0.0465	0.0536	-0.0197	0.9075
perfimpt	0.1111	-0.0810	-0.0756	0.0256	-0.8836
envrminr	-0.1734	-0.0349	0.0855	-0.8666	-0.0136
shdcarpl	0.1636	-0.0277	-0.0551	0.8665	0.0775
lk4whldr	0.1676	0.0247	0.0318	0.1041	0.8556
nordtrps	-0.0637	-0.1018	-0.8665	-0.0135	-0.0410
carefmny	-0.7634	-0.1452	-0.1975	-0.3081	-0.0805
miniboxy	0.1229	0.8419	-0.1072	0.0511	0.0131
suvcmpct	0.0838	0.8191	0.2022	0.0371	-0.0036
kidsbulk	0.1799	0.0173	0.8248	0.0573	0.0225
wntguzlr	-0.3571	0.0323	-0.0074	-0.7628	0.0150
buyhghnd	0.8147	0.1758	0.0222	0.0541	0.0981
tkvacatn	0.6539	-0.0310	0.2570	0.4583	0.0225
homlrgst	0.3309	-0.6794	0.1535	0.3166	0.0855
aftrschl	0.1951	-0.1092	0.7754	-0.1132	0.1819
strngwrn	0.2730	-0.2587	0.0828	0.0609	0.7354
lthrbetr	0.7107	-0.1865	0.2461	0.2897	0.0675
noparkrm	0.1739	0.8066	0.0121	-0.0872	-0.0164
accesfun	0.6784	-0.0387	0.3008	0.3695	0.0025
pricqual	0.7823	-0.1904	-0.0821	-0.1353	0.0041
passnimp	-0.6475	-0.0197	-0.3995	-0.2767	0.0143
next2str	0.2581	-0.7429	0.1056	-0.1152	-0.0068
stylclth	0.6036	0.2363	0.1824	0.4262	-0.0270
twoincom	0.7562	0.1170	-0.0943	-0.0741	0.0951
secbiggr	-0.0759	0.7593	0.0587	0.0332	-0.0825
needbetw	0.1264	0.7575	-0.0111	0.0434	0.0424
prmsound	0.6819	-0.0175	0.1678	0.2897	0.0731
imprtapp	0.5094	-0.0102	0.3456	0.3495	0.2000
nohummer	0.0595	0.7064	0.0455	-0.0376	0.0384

- Factor 3 has high **negative correlation** with nordtrps (We don"t go on road trips with the family) and high **positive correlation** with kidtrans (We need a car that helps transport our kids and their friends), kidsbulk (Our kids tend to take a lot of bulky items and toys with them) and aftrschl (We engage in more after-school activities than most families)

- Factor 3 is related to **kids' needs for vehicle.**

Factor Analysis - RC4

	RC1	RC2	RC3	RC5	RC4
kidtrans	0.1237	0.0023	0.9330	-0.0196	0.0067
safeimpt	0.0261	0.0465	0.0536	-0.0197	0.9075
perfimpt	0.1111	-0.0810	-0.0756	0.0256	-0.8836
envrminr	-0.1734	-0.0349	0.0855	-0.8666	-0.0136
shdcarpl	0.1636	-0.0277	-0.0551	0.8665	0.0775
lk4whldr	0.1676	0.0247	0.0318	0.1041	0.8556
nordtrps	-0.0637	-0.1018	-0.8665	-0.0135	-0.0410
carefmny	-0.7634	-0.1452	-0.1975	-0.3081	-0.0805
miniboxy	0.1229	0.8419	-0.1072	0.0511	0.0131
suvcmpct	0.0838	0.8191	0.2022	0.0371	-0.0036
kidsbulk	0.1799	0.0173	0.8248	0.0573	0.0225
wntguzlr	-0.3571	0.0323	-0.0074	-0.7628	0.0150
buyhghnd	0.8147	0.1758	0.0222	0.0541	0.0981
tkvacatn	0.6539	-0.0310	0.2570	0.4583	0.0225
homlrgst	0.3309	-0.6794	0.1535	0.3166	0.0855
aftrschl	0.1951	-0.1092	0.7754	-0.1132	0.1819
strngwrn	0.2730	-0.2587	0.0828	0.0609	0.7354
lthrbetr	0.7107	-0.1865	0.2461	0.2897	0.0675
noparkrm	0.1739	0.8066	0.0121	-0.0872	-0.0164
accesfun	0.6784	-0.0387	0.3008	0.3695	0.0025
pricqual	0.7823	-0.1904	-0.0821	-0.1353	0.0041
passnimp	-0.6475	-0.0197	-0.3995	-0.2767	0.0143
next2str	0.2581	-0.7429	0.1056	-0.1152	-0.0068
stylclth	0.6036	0.2363	0.1824	0.4262	-0.0270
twoincom	0.7562	0.1170	-0.0943	-0.0741	0.0951
secbiggr	-0.0759	0.7593	0.0587	0.0332	-0.0825
needbetw	0.1264	0.7575	-0.0111	0.0434	0.0424
prmsound	0.6819	-0.0175	0.1678	0.2897	0.0731
imprtapp	0.5094	-0.0102	0.3456	0.3495	0.2000
nohummer	0.0595	0.7064	0.0455	-0.0376	0.0384

- Factor 4 has high **negative correlation** with perfimpt (Performance is very important in a car), and high **positive correlation** with safeimpt (Auto safety is very important to me), lk4whldr (Four-wheel drive is a very attractive option) and strngwrn (Warranty protection needs to be strong on a new car).

- Factor 4 is related to **safety**

Factor Analysis - RC5

	RC1	RC2	RC3	RC5	RC4
kidtrans	0.1237	0.0023	0.9330	-0.0196	0.0067
safeimpt	0.0261	0.0465	0.0536	-0.0197	0.9075
perfimpt	0.1111	-0.0810	-0.0756	0.0256	-0.8836
envrminr	-0.1734	-0.0349	0.0855	-0.8666	-0.0136
shdcarpl	0.1636	-0.0277	-0.0551	0.8665	0.0775
lk4whldr	0.1676	0.0247	0.0318	0.1041	0.8556
nordtrps	-0.0637	-0.1018	-0.8665	-0.0135	-0.0410
carefmny	-0.7634	-0.1452	-0.1975	-0.3081	-0.0805
miniboxy	0.1229	0.8419	-0.1072	0.0511	0.0131
suvcmpct	0.0838	0.8191	0.2022	0.0371	-0.0036
kidsbulk	0.1799	0.0173	0.8248	0.0573	0.0225
wntguzlr	-0.3571	0.0323	-0.0074	-0.7628	0.0150
buyhghnd	0.8147	0.1758	0.0222	0.0541	0.0981
tkvacatn	0.6539	-0.0310	0.2570	0.4583	0.0225
homlrgst	0.3309	-0.6794	0.1535	0.3166	0.0855
aftrschl	0.1951	-0.1092	0.7754	-0.1132	0.1819
strngwrn	0.2730	-0.2587	0.0828	0.0609	0.7354
lthrbetr	0.7107	-0.1865	0.2461	0.2897	0.0675
noparkrm	0.1739	0.8066	0.0121	-0.0872	-0.0164
accesfun	0.6784	-0.0387	0.3008	0.3695	0.0025
pricqual	0.7823	-0.1904	-0.0821	-0.1353	0.0041
passnimp	-0.6475	-0.0197	-0.3995	-0.2767	0.0143
next2str	0.2581	-0.7429	0.1056	-0.1152	-0.0068
stylclth	0.6036	0.2363	0.1824	0.4262	-0.0270
twoincom	0.7562	0.1170	-0.0943	-0.0741	0.0951
secbiggr	-0.0759	0.7593	0.0587	0.0332	-0.0825
needbetw	0.1264	0.7575	-0.0111	0.0434	0.0424
prmsound	0.6819	-0.0175	0.1678	0.2897	0.0731
imprtapp	0.5094	-0.0102	0.3456	0.3495	0.2000
nohummer	0.0595	0.7064	0.0455	-0.0376	0.0384

- Factor 5 has high **negative correlation** with envrminr (The environmental impact of automobiles is relatively minor) and wntguzlr (Will buy what I want even if it is a “gas guzzler”), and high **positive correlation** with shdcarpl (Everyone should carpool or take public transportation).

- Factor 5 is related to **environmental impact.**

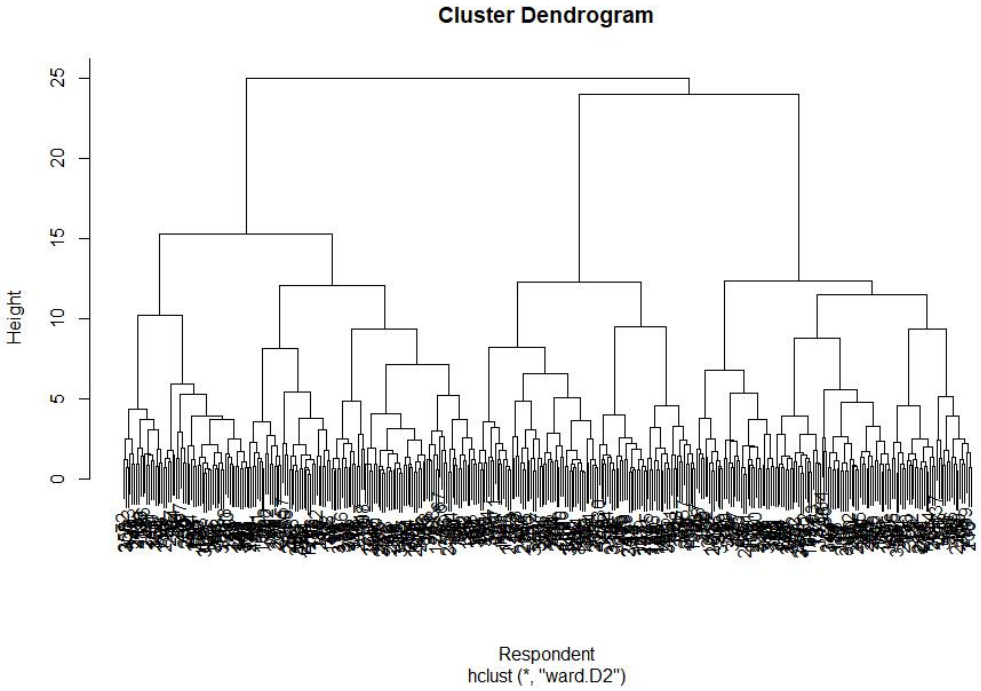
Regression using Factor Scores

```
Call:
lm(formula = mvliking ~ RC1 + RC2 + RC3 + RC4 + RC5, data = factor_scores)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9530 -1.5723 -0.0992  1.6137  6.1489

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.8425     0.1104  43.878  < 2e-16 ***
RC1           1.0439     0.1105   9.447  < 2e-16 ***
RC2           0.9854     0.1105   8.918  < 2e-16 ***
RC3           0.1838     0.1105   1.664    0.097 .
RC4          -0.5570     0.1105  -5.040  7.1e-07 ***
RC5          -0.1418     0.1105  -1.284    0.200
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.207 on 394 degrees of freedom
Multiple R-squared:  0.3351,    Adjusted R-squared:  0.3267
F-statistic: 39.72 on 5 and 394 DF,  p-value: < 2.2e-16
```

- Despite that the initial model (slide 4) fit is significant, there are too many predictors in the model and several of the coefficients are non-significant, indicating that these variables do not contribute much to the model.

- For the factors regression model, the model fit, p-value, is the same as for the model using all the original predictors, as expected.
  - Two of the 5 factors (3 and 5) are insignificant. If we drop them and run the regression again using only factors 1, 2, and 4 we get a regression with all factors being significant.
  - Since the variability of the coefficient estimates are not greatly affected by the collinearity issue, the coefficient estimates are now larger than their respective standard errors compared to the original model with all the predictors.

Market Segmentation - Hierarchical Cluster



The dendrogram shows that it is more appropriate to divide the respondents in the dataset into three clusters. Going lower down the hierarchy will be harder to manage.

Market Segmentation - k-Means Clustering

Cluster Centers:

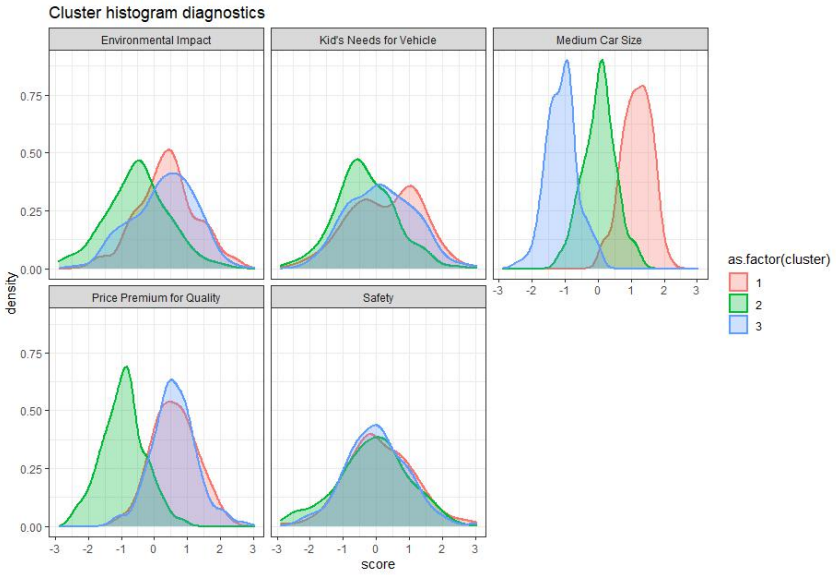
	RC1	RC2	RC3	RC5	RC4
1	0.6107503	1.17402497	0.2725255	0.3807394	0.11476200
2	-0.9323024	0.04589391	-0.3249301	-0.5114655	-0.09657823
3	0.5949299	-1.09336582	0.1546210	0.2860140	0.01612336

- We will focus on 3 clusters for market positioning

Cluster	What Factor Scores They Differ	Label	Total Respondents
1	Strong positive association with RC2	Comfortable and spacious car lovers	114
2	Strong negative association with RC1	Economical car buyers	157
3	Strong negative association with RC2	Compact and mini car lovers	129

Concept Preference by Segments

Cluster Diagnostics on RC1-5



Regression of mvliking on cluster id categorical variable

```
Call:
lm(formula = mvliking ~ as.factor(cluster), data = factor_scores)

Residuals:
    Min      1Q  Median      3Q     Max
-5.6053 -1.8981  0.1019  2.1019  5.1019

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)               6.6053     0.2289  28.857  < 2e-16 ***
as.factor(cluster)2      -2.7072     0.3007  -9.002  < 2e-16 ***
as.factor(cluster)3      -2.1712     0.3142  -6.911 1.93e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.444 on 397 degrees of freedom
Multiple R-squared:  0.1787,    Adjusted R-squared:  0.1745
F-statistic: 43.18 on 2 and 397 DF,  p-value: < 2.2e-16
```

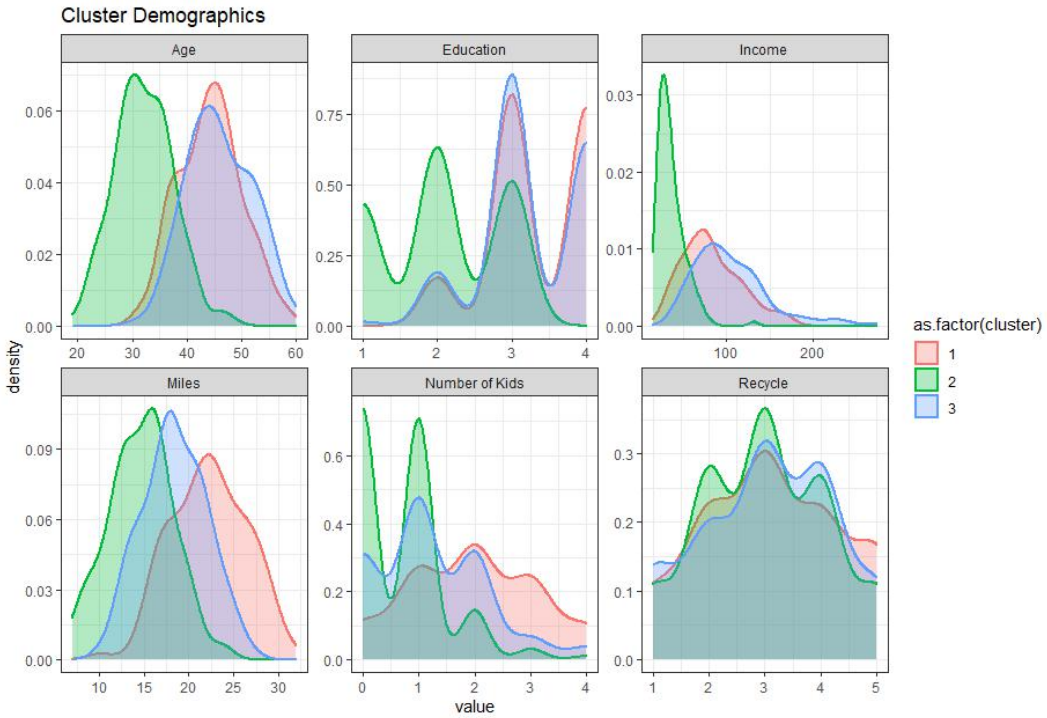
Cross tabulation of cluster membership and discrete levels of mvliking

Cluster - mvliking crosstable

cluster	*mvliking* 1	2	3	4	5	6	7	8	9	***Total***
1	5 4.4 %	3 2.6 %	5 4.4 %	6 5.3 %	15 13.2 %	16 14 %	12 10.5 %	20 17.5 %	32 28.1 %	114 100 %
2	39 24.8 %	15 9.6 %	20 12.7 %	26 16.6 %	16 10.2 %	15 9.6 %	10 6.4 %	5 3.2 %	11 7 %	157 100 %
3	25 19.4 %	8 6.2 %	17 13.2 %	17 13.2 %	21 16.3 %	11 8.5 %	11 8.5 %	7 5.4 %	12 9.3 %	129 100 %
Total	69 17.2 %	26 6.5 %	42 10.5 %	49 12.2 %	52 13 %	42 10.5 %	33 8.2 %	32 8 %	55 13.8 %	400 100 %

χ^2=82.160 · df=16 · Cramer's V=0.320 · p=0.000

Demographic by Segments



*Note: Summary table is available in Appendix.

- Based on cluster centers, cluster 1 tends to value comfortable and spacious car. This is reasonable because they have the highest annual mileage driven, and they have an average of 2 children. This cluster also has relatively higher income.
- The cluster centers in slide 14 show that cluster 2 values affordability. This is aligned with the demographic profile since majority of them have lower income. This cluster also tends to be younger. They drive the lowest mileage and mostly have 0 or 1 child.
- Cluster 3 is the oldest among the three clusters, with the highest median income (see Appendix). Fifty-five percent of them are female, and they have an average of 1 child.

Recommendations

- We Recommend to Target on Cluster 1
  - From the results of regression and crosstab:
    - Cluster 1 has the highest liking score for microvans
  - From the results of factor analysis and clustering:
    - Cluster 1 has needs for bigger car space
  - From the results of demographics:
    - Cluster 1 has medium income level, high annual mileage, average of 2 kids
- Positioning
  - Big (small to large family with kids) and durable (high average mileage per year)
  - Packed with features for kids - in-car AV system, cup holders, kids' safety features, etc
  - Good high end visual quality, leather trim options must present
  - Medium and affordable price point

Assumptions and Limitations

Assumptions:

- We assume the GPA wants the most cost-saving and manageable segmentation, so we did analysis for 3 clusters instead of 4

Limitations:

- The study did not consider how much respondents are willing to pay for a microvan
- 4 clusters would be better in more granularly targeting the customers
- Small sample size can be a concern since the usable dataset consists only 400 observations

Next Steps:

- Conduct conjoint study to test different price points
- Analyze competitors and determine precise marketing position
- Draft marketing mix strategies (4Ps)

Research Design Improvement

There are other customer data we want to collect to better target at and position on microvan, including:

- **Preference on electric, hybrid or gas fuel:** to include more specific environment features that customers care
- **Perception on more car types** such as hatchback, pick-up trucks: to understand what features the customers desire from each car type
- Create **perceptual map** and use the “white space”: to discover new customer segmentations
- **Budget and buying intent**: to finalize price tiers and design more reasonable add-on packages

Appendix

Variable Key

v01	kidtrans	We need a car that helps transport our kids and their friends.
v02	miniboxy	Current minivans are simply too boxy and large.
v03	lthrbetr	Leather seats are dramatically better than cloth.
v04	secbiggr	If we got a second car, it would need to be bigger than a standard sedan.
v05	safeimpt	Auto safety is very important to me.
v06	buyhghnd	We tend to buy higher-end cars.
v07	pricqual	Car prices strongly reflect underlying production quality.
v08	prmsound	A premium sound and entertainment system helps on long car trips.
v09	perfimpt	Performance is very important in a car.
v10	tkvacatn	We try to take as many vacations as possible.
v11	noparkrm	Our current residence doesn't have a lot of parking room.
v12	homlrgst	Our home is among the largest in the neighborhood.
v13	envrminr	The environmental impact of automobiles is relatively minor.
v14	needbetw	There needs to be something between a sedan and a minivan.
v15	suvcmpct	I like SUVs more than minivans since they're more compact.
v16	next2str	My next car will be a two-seater.
v17	carefmny	We are careful with money.
v18	shdcarpl	I think everyone should carpool or take public transportation.
v19	imprtapp	Most of our appliances are imported.
v20	lk4whldr	Four-wheel drive is a very attractive option.
v21	kidsbulk	Our kids tend to take a lot of bulky items and toys with them.
v22	wntguzlr	I will buy what I want even if it is a “gas guzzler”.
v23	nordtrps	We don’t go on road trips with the family
v24	stylclth	We tend to purchase stylish clothes for the family.
v25	strngwrn	Warranty protection needs to be strong on a new car.
v26	passnimp	Passion for one’s job is more important than pay.
v27	twoincom	Our family would find it hard to subsist on just one income.
v28	nohummer	I am not interested in owning a vehicle like a Hummer.
v29	aftrschl	We engage in more after-school activities than most families.
v30	accesfun	Accessories really make a car more fun to drive.

age	Age of respondent in years
income	Annual household income in thousands of dollars
miles	Total annual amount driven by household members in thousands of miles
numkids	Number of children (aged 0-18) residing in household
female	Whether or not the respondent is a female
educ	Education level of respondent (1 = High School, 2 = Some College, 3 = Undergraduate Degree, 4 = Graduate Degree)
recycle	Self-reported recycling compared to average (1 = Much Less, 2 = Somewhat Less, 3 = Average, 4 = Somewhat More, 5 = Much More)

Demographic Summary Table by Clusters

	Cluster 1 Comfortable & spacious car lovers	Cluster 2 Economical car buyers	Cluster 3 Compact & mini car lovers
Number of respondents	114	157	129
Mean Age	44.3	32.1	46.0
Median Income (in thousands)	77.5	32.0	96.0
Mean annual miles driven (in thousands)	22.3	14.7	18.4
Mean Number of Kids	2.0	0.7	1.2
Proportion of Female	59.7%	49.0%	55.0%
Median Education Level	Undergraduate Degree	Some College	Undergraduate Degree
Mean Recycling Score	3.1	3.0	3.0