

Exercise: Bank Marketing dataset**Variables inspection**

Continuous variables are *log10* transformed (with some transformations adding an offset to avoid NAs), scaled and centered.

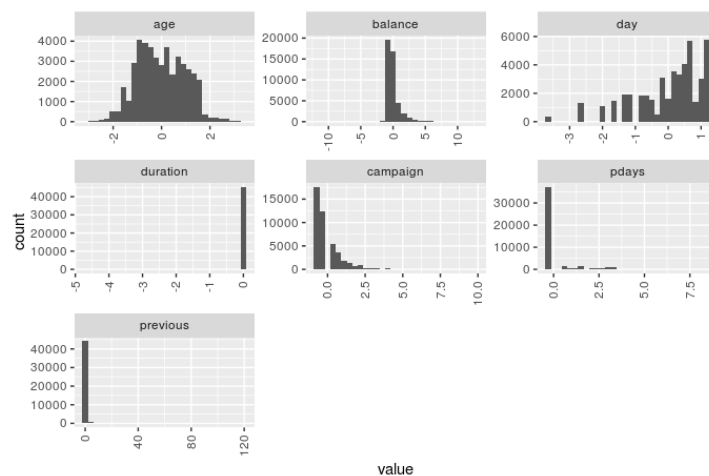


Figure 1 - continuous variables inspection, after \log_{10} transform

R will create dummy variables automatically when fitting models:

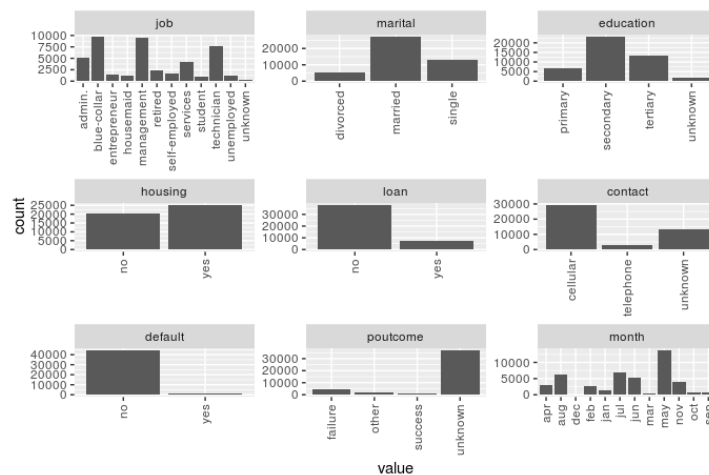


Figure 2 - categorical variables inspection

No special correlation is found, then models using all categorical and continuous variables are generated.

2/3 of data is being used as *learn*, while the remaining 1/3 as *test* for all the models.

Model fitting

<i>Algorithm</i>	<i>Accuracy</i>
LDA	86.34%
LDA_CV	86.68%
RDA	85.90%
RDA_CV	85.75%
QDA	85.89%
Logistic regression	<i>Learning data: 9.55% ERROR</i> <i>Test data: 9.78% ERROR</i>

Results look nice, being the models accurate ~86% of time, which is pretty good.

Algorithm	Confusion Matrix																														
LDA	<table><tr><td></td><td colspan="2">Reference</td></tr><tr><td>Prediction</td><td>no</td><td>yes</td></tr><tr><td>no</td><td>12202</td><td>920</td></tr><tr><td>yes</td><td>1120</td><td>828</td></tr></table>		Reference		Prediction	no	yes	no	12202	920	yes	1120	828																		
	Reference																														
Prediction	no	yes																													
no	12202	920																													
yes	1120	828																													
LDA_CV	<table><tr><td></td><td colspan="2">Reference</td></tr><tr><td>Prediction</td><td>no</td><td>yes</td></tr><tr><td>no</td><td>23518</td><td>932</td></tr><tr><td>yes</td><td>3082</td><td>2609</td></tr></table>		Reference		Prediction	no	yes	no	23518	932	yes	3082	2609																		
	Reference																														
Prediction	no	yes																													
no	23518	932																													
yes	3082	2609																													
QDA	<table><tr><td></td><td colspan="2">Reference</td></tr><tr><td>Prediction</td><td>no</td><td>yes</td></tr><tr><td>no</td><td>12074</td><td>877</td></tr><tr><td>yes</td><td>1248</td><td>871</td></tr></table>		Reference		Prediction	no	yes	no	12074	877	yes	1248	871																		
	Reference																														
Prediction	no	yes																													
no	12074	877																													
yes	1248	871																													
QDA_CV	<table><tr><td></td><td colspan="2">Reference</td></tr><tr><td>Prediction</td><td>no</td><td>yes</td></tr><tr><td>no</td><td>24146</td><td>1840</td></tr><tr><td>yes</td><td>2454</td><td>1701</td></tr></table>		Reference		Prediction	no	yes	no	24146	1840	yes	2454	1701																		
	Reference																														
Prediction	no	yes																													
no	24146	1840																													
yes	2454	1701																													
RDA	<table><tr><td></td><td colspan="2">Reference</td></tr><tr><td>Prediction</td><td>no</td><td>yes</td></tr><tr><td>no</td><td>12083</td><td>888</td></tr><tr><td>yes</td><td>1239</td><td>860</td></tr></table>		Reference		Prediction	no	yes	no	12083	888	yes	1239	860																		
	Reference																														
Prediction	no	yes																													
no	12083	888																													
yes	1239	860																													
Logistic regression	<table><tr><td></td><td colspan="2">Learning data:</td></tr><tr><td></td><td colspan="2">Reference</td></tr><tr><td>Prediction</td><td>no</td><td>yes</td></tr><tr><td>no</td><td>25943</td><td>2223</td></tr><tr><td>yes</td><td>657</td><td>1318</td></tr><tr><td></td><td colspan="2">Test data:</td></tr><tr><td></td><td colspan="2">Reference</td></tr><tr><td>Prediction</td><td>no</td><td>yes</td></tr><tr><td>no</td><td>12975</td><td>1127</td></tr><tr><td>yes</td><td>347</td><td>621</td></tr></table>		Learning data:			Reference		Prediction	no	yes	no	25943	2223	yes	657	1318		Test data:			Reference		Prediction	no	yes	no	12975	1127	yes	347	621
	Learning data:																														
	Reference																														
Prediction	no	yes																													
no	25943	2223																													
yes	657	1318																													
	Test data:																														
	Reference																														
Prediction	no	yes																													
no	12975	1127																													
yes	347	621																													