

In [14]:

```
%matplotlib inline
import pandas
from sklearn import linear_model as lm
from plotnine import *
import math

root = '/Users/rix0rrr/Google Drive/Hackathon 28_4'
filename = root + '/Team 3/Input voor de lineare regressie-v5.csv'
join_file = root + '/Team 3/Input voor de lineare regressie-financieel-v3.csv'

#df = pandas.read_csv(filename, index_col='BEVOEGD_GEZAGNAAM', encoding='utf-8')
df = pandas.read_csv(filename, index_col='Key', encoding='utf-8')
joinerts = pandas.read_csv(join_file, index_col='Key', encoding='utf-8')

joinert_cols = list(joinerts.columns.values)

df = df.join(joinerts)

for col in joinert_cols:
    try:
        df[col + ' per leerling'] = df[col] / df['AANTAL_LEERLINGEN']
    except Exception as e:
        print(e)
        pass

df = df.drop(joinert_cols, axis=1)
cols = list(df.columns.values)

# Ons jaar
df = df.loc[df['JAAR'] == 2016]

# geen kleine scholen
df = df.loc[df['OMVANG_FORMATIE_TOTAAL'] >= 10]

# omvang besturen
#df = df.loc[df['Aantal instellingen'] >= 6]

print("Besturen in analyse: %s" % len(df))

df = df.sort_values('AANDEEL_FORMATIE_OP_LB')
df.loc[:, 'cat'] = 'overig'
df.loc[:50, 'cat'] = 'slechtst'
df.loc[-50:, 'cat'] = 'best'

print('De allerbeste is %s dus die doet niet meer mee' % df[-1:][['BEVOEGD_GEZAGN'])

df = df[:-1]

cols = [c for c in cols if not c.startswith('AANDEEL_FORMATIE_')]

graphs = []
```

```

for x_as in cols:
    #x_as = 'SolvabiliteitI.2016'
    y_as = 'AANDEEL_FORMATIE_OP_LB'

    rows = df[[x_as, y_as, 'cat']].dropna()
    if not len(rows):
        print('NO NON-NA ROWS')
        continue

    try:
        best_corr = rows.loc[rows['cat'] == 'best', [x_as, y_as]].corr()[x_as][y_as]
        orig_corr = rows.loc[rows['cat'] == 'orig', [x_as, y_as]].corr()[x_as][y_as]
        slechtst_corr = rows.loc[rows['cat'] == 'slechtst', [x_as, y_as]].corr()[x_as][y_as]
        alles_corr = rows.loc[:, [x_as, y_as]].corr()[x_as][y_as]

        corrs = [best_corr, orig_corr, slechtst_corr, alles_corr]

        corrs = [abs(c) for c in corrs if not math.isnan(c)]
        max_corr = max(corrs) if corrs else 0

        graphs.append((max_corr, ggplot(rows, aes(x_as, y_as, color='factor(cat)'))
            + ggtile('%s (%s rows, %.2f/%.2f/%.2f alles %.2f)' % (x_as, len(rows),
            + geom_point()
            + stat_smooth(method='lm'))))
    except Exception as e:
        print(e)
        pass

graphs.sort(key=lambda x: x[0], reverse=True)
graphs = [x[1] for x in graphs]
graphs

```

ufunc 'true\_divide' not supported for the input types, and the input  
s could not be safely coerced to any supported types according to th  
e casting rule ''safe''

ufunc 'true\_divide' not supported for the input types, and the input  
s could not be safely coerced to any supported types according to th  
e casting rule ''safe''

ufunc 'true\_divide' not supported for the input types, and the input  
s could not be safely coerced to any supported types according to th  
e casting rule ''safe''

ufunc 'true\_divide' not supported for the input types, and the input  
s could not be safely coerced to any supported types according to th  
e casting rule ''safe''

ufunc 'true\_divide' not supported for the input types, and the input  
s could not be safely coerced to any supported types according to th  
e casting rule ''safe''

ufunc 'true\_divide' not supported for the input types, and the input  
s could not be safely coerced to any supported types according to th  
e casting rule ''safe''

Besturen in analyse: 805

De allerbeste is  
 \_OP\_LB

BEVOEGD\_GEZAGNAAM AANDEEL\_FORMATIE

Key

42679-2016 Stichting Kronenburgh  
et niet meer mee

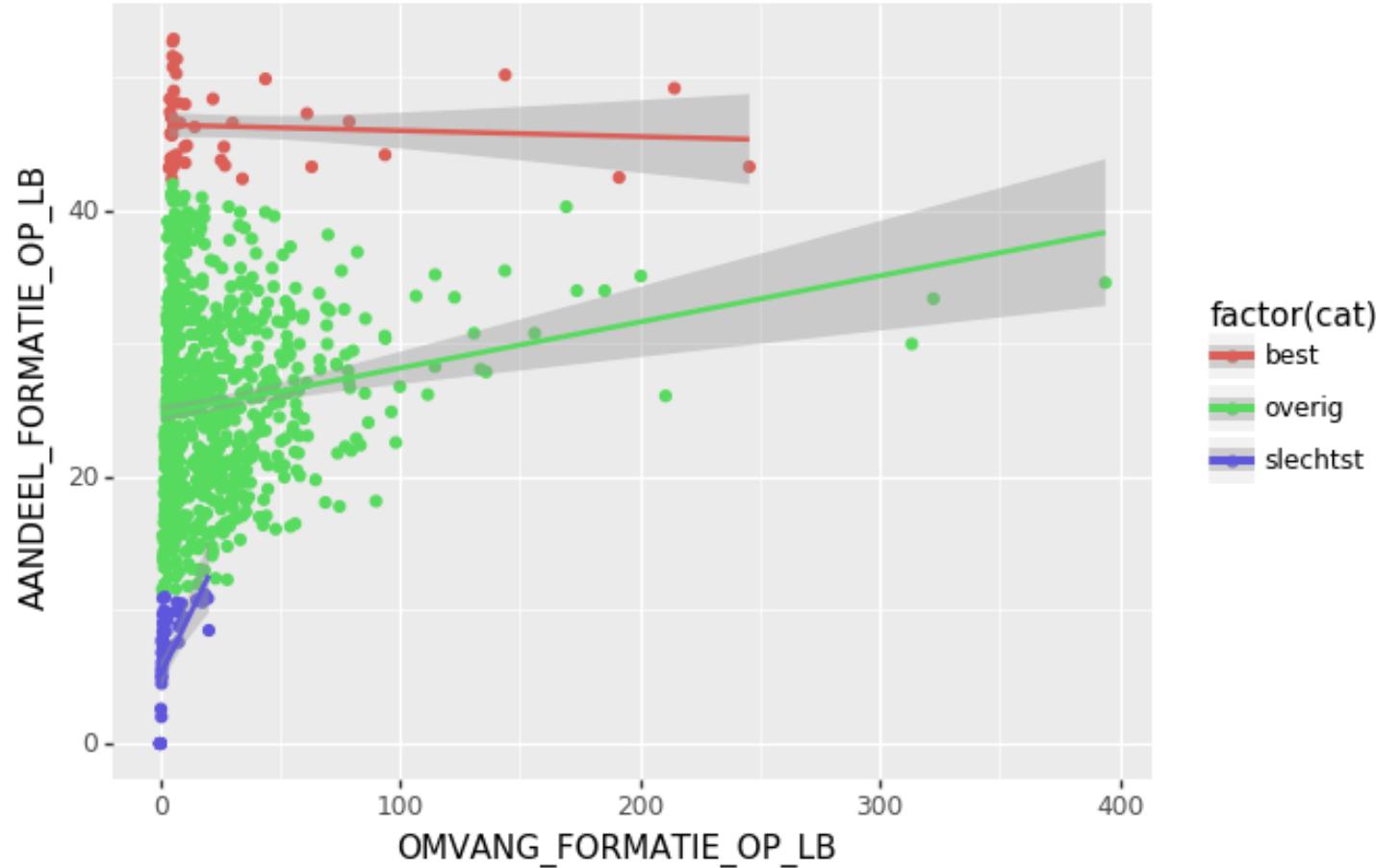
#### 74.4 dus die do

## 'BEVOEGD GEZAGNAAM'

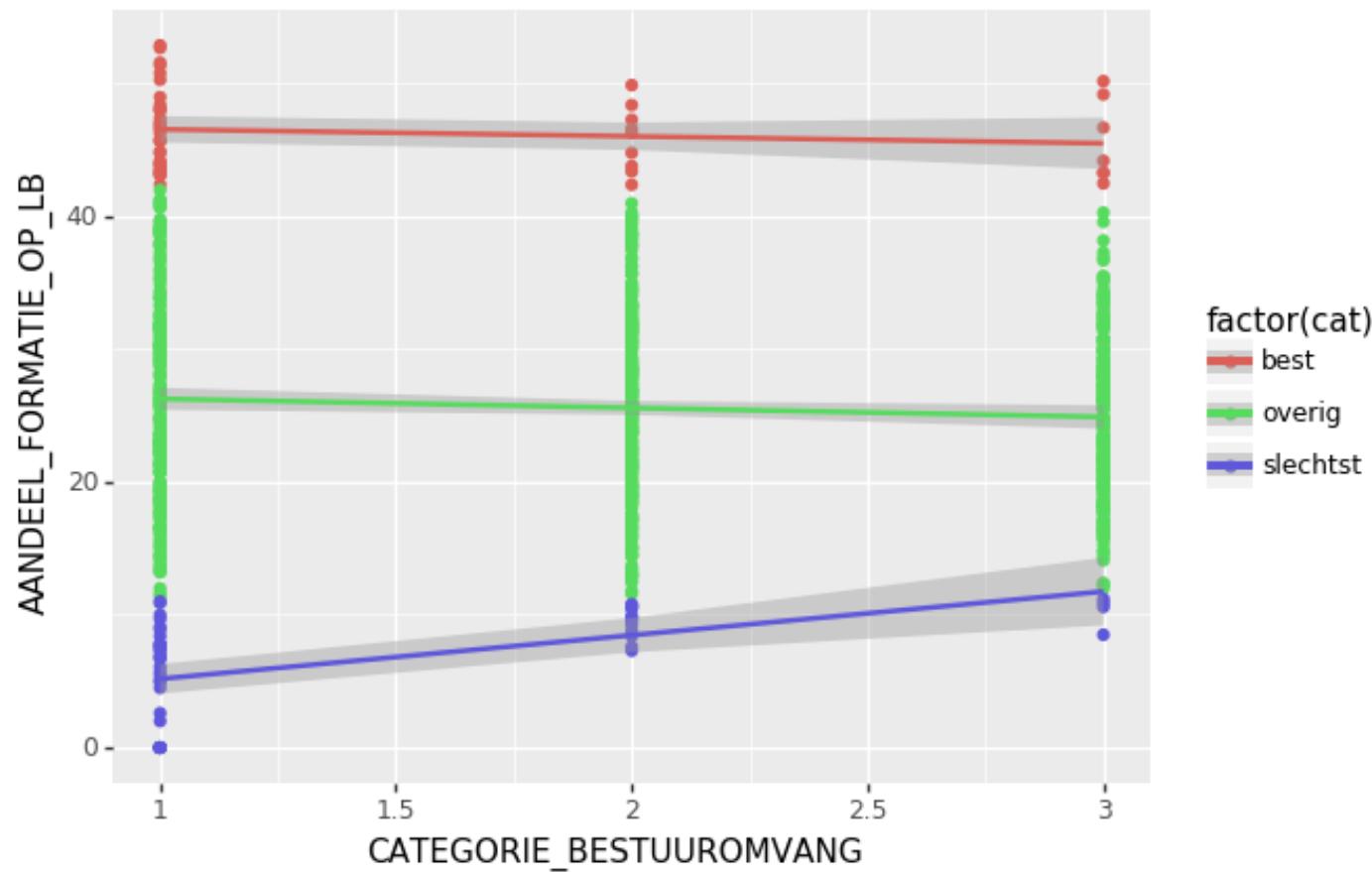
## NO NON-NA ROWS

OMVANG FORM

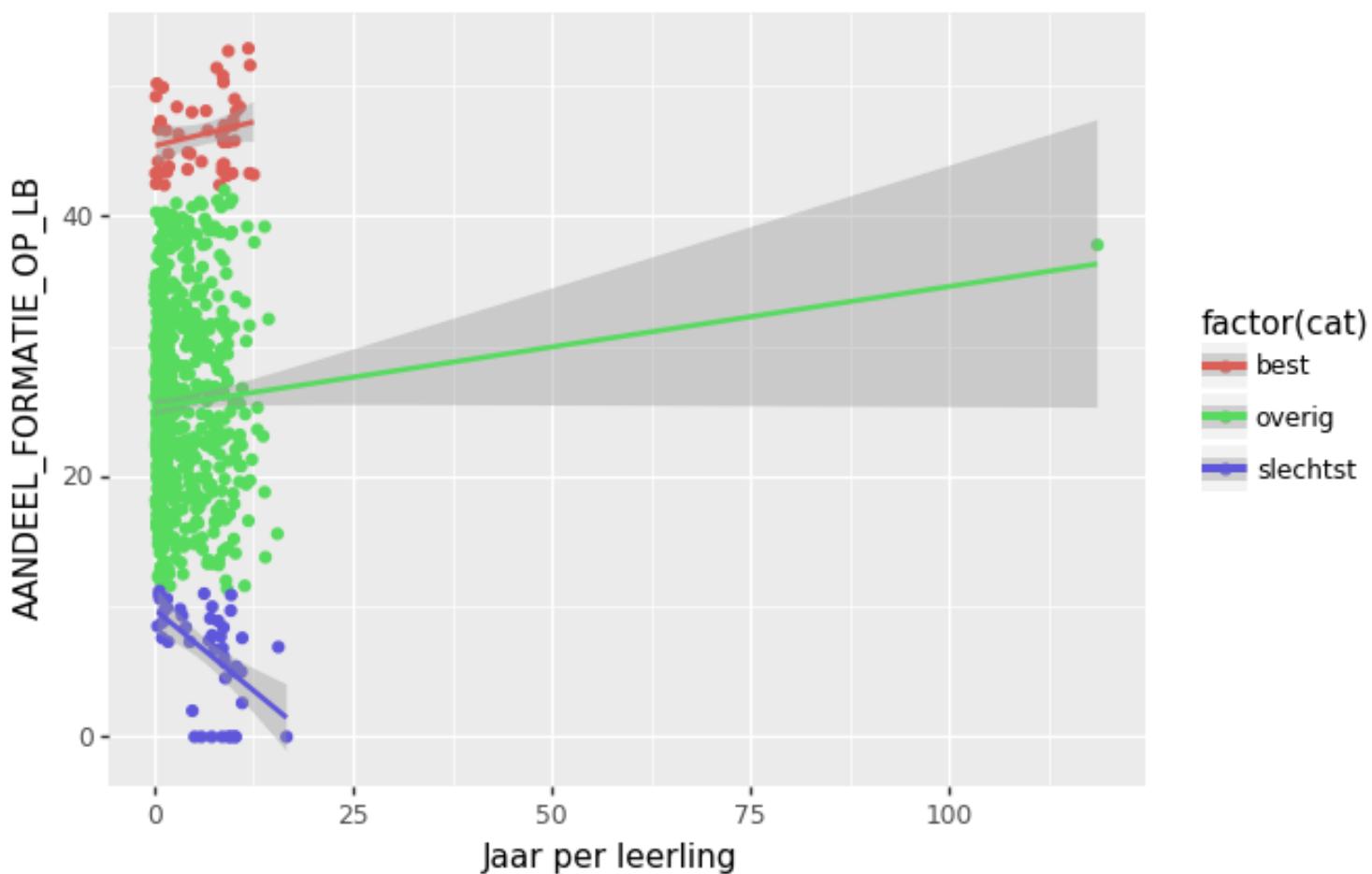
OMVANG\_FORMATIE\_OP\_LB (804 rows, -0.09/0.17/0.56 alles 0.20)



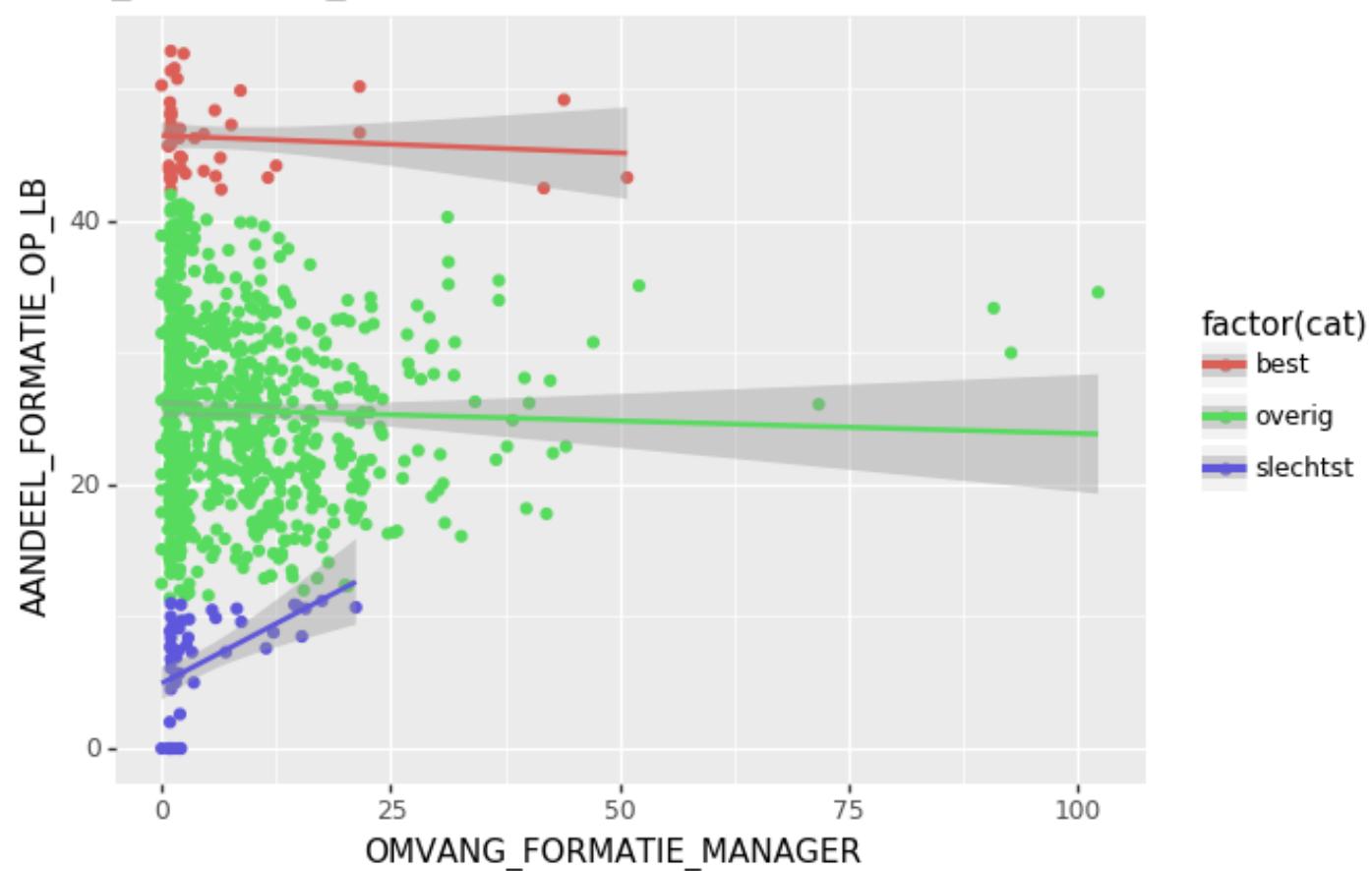
CATEGORIE\_BESTUUROMVANG (804 rows, -0.13/-0.07/0.54 alles -0.03)



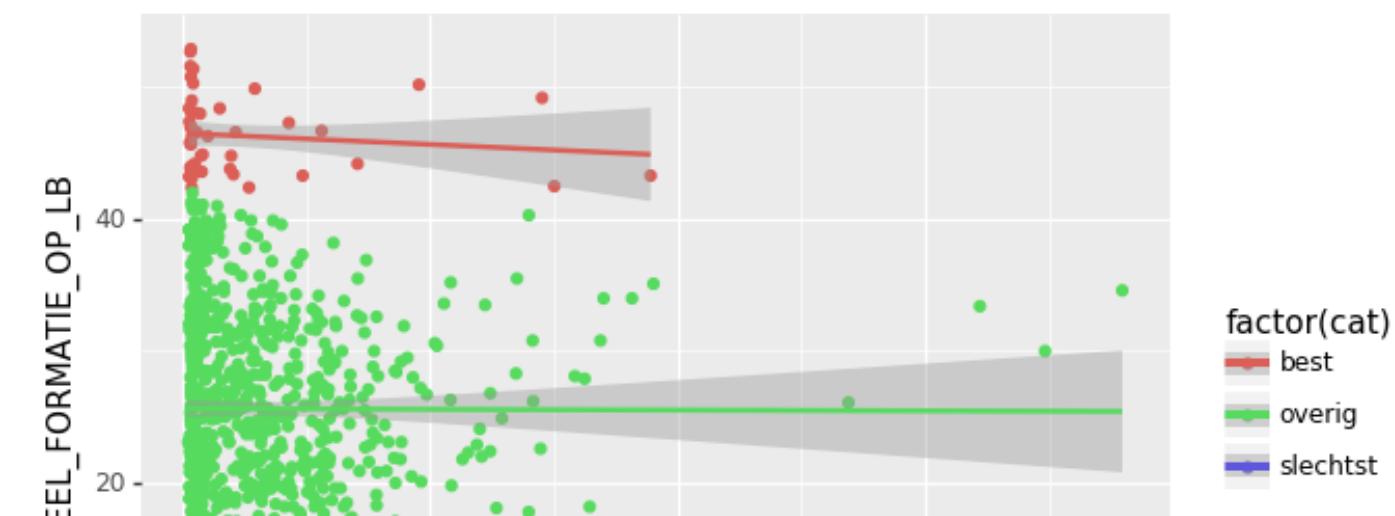
Jaar per leerling (791 rows, 0.20/0.07/-0.54 alles 0.04)

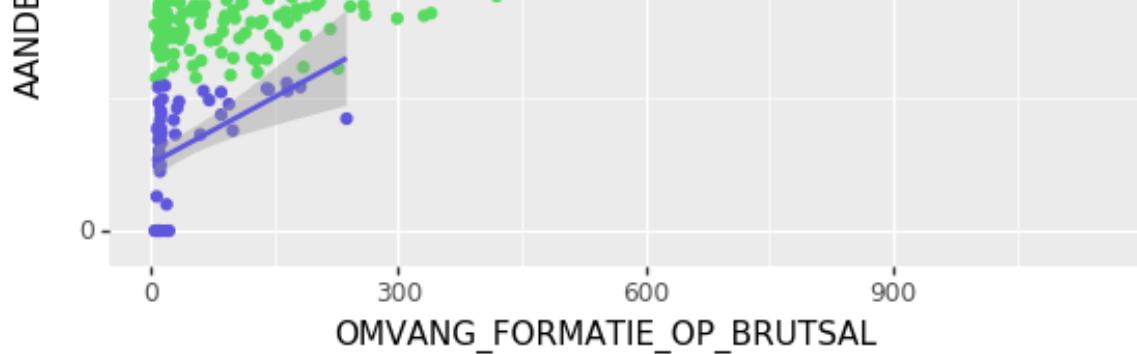


OMVANG\_FORMATIE\_MANAGER (804 rows, -0.10/-0.03/0.50 alles 0.00)

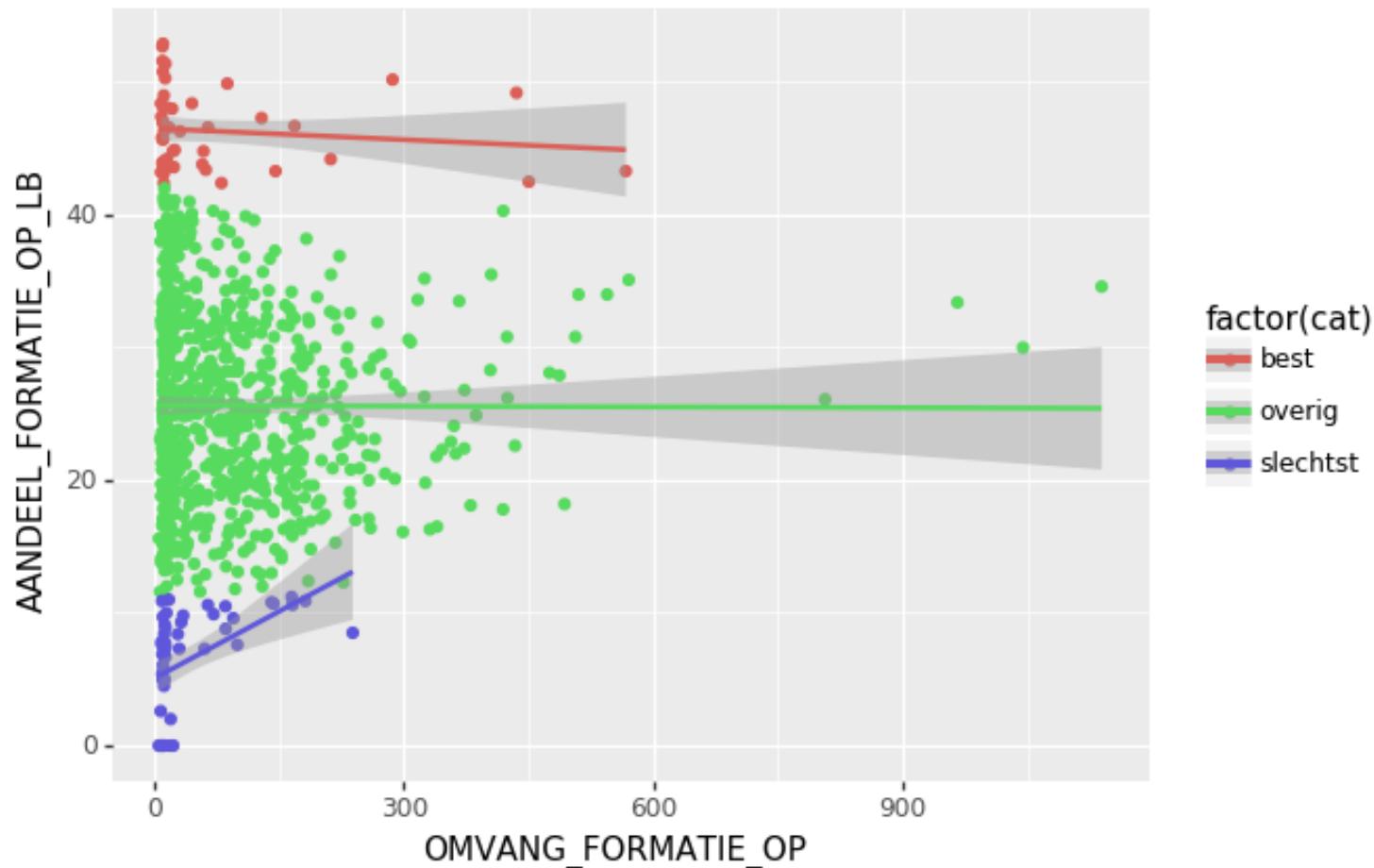


OMVANG\_FORMATIE\_OP\_BRUTSAL (804 rows, -0.12/-0.00/0.48 alles 0.03)

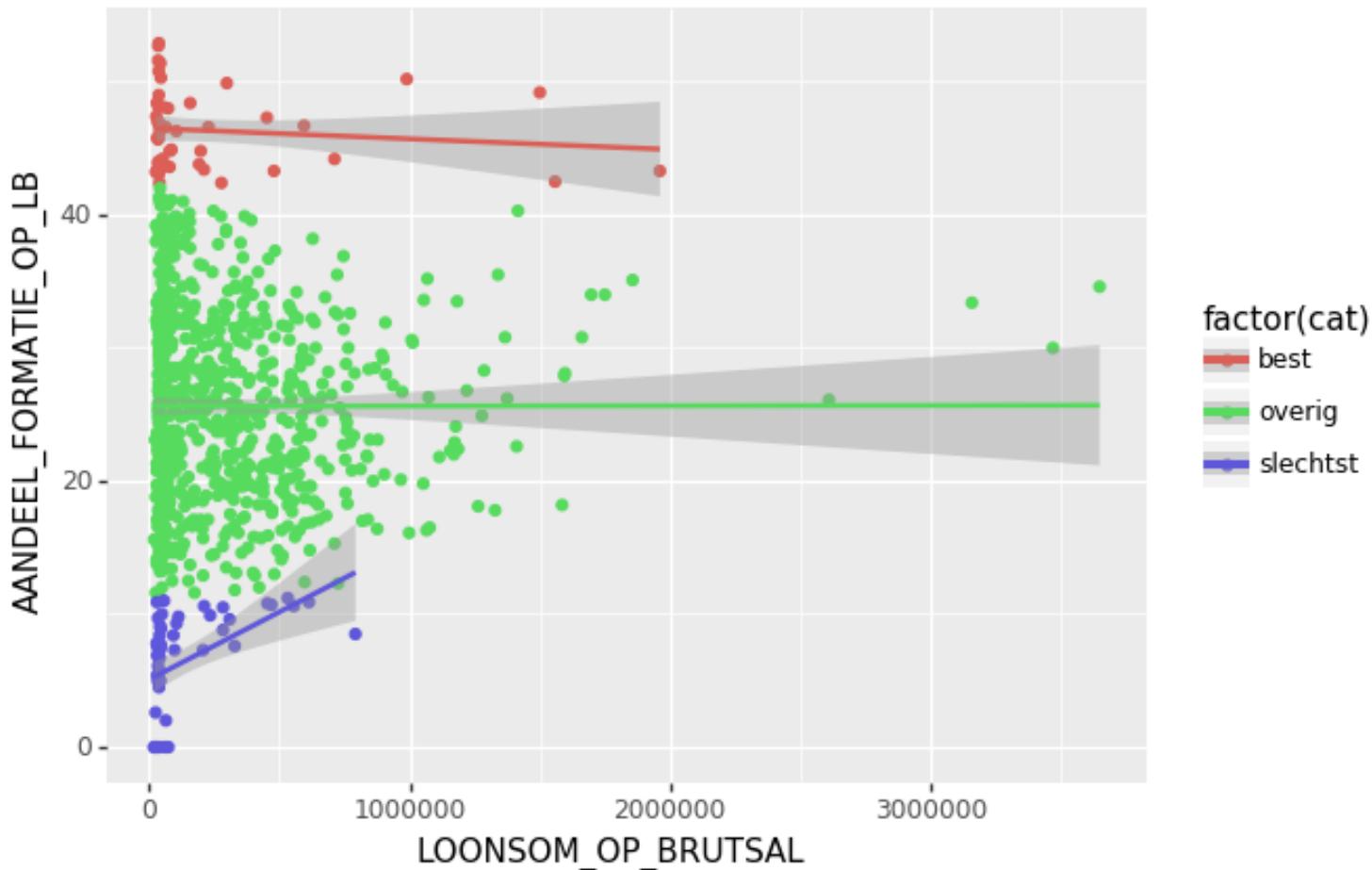




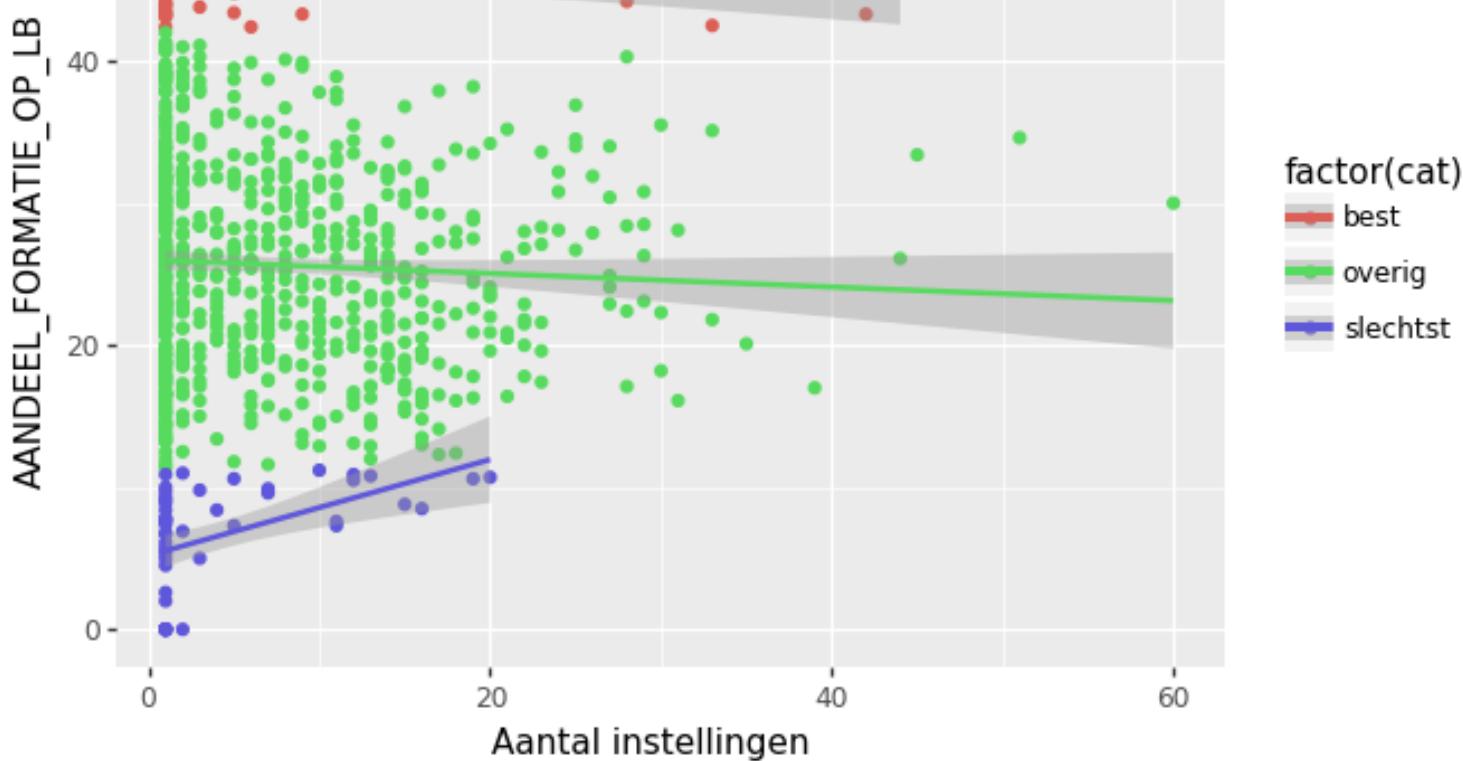
OMVANG\_FORMATIE\_OP (804 rows, -0.12/-0.00/0.48 alles 0.03)



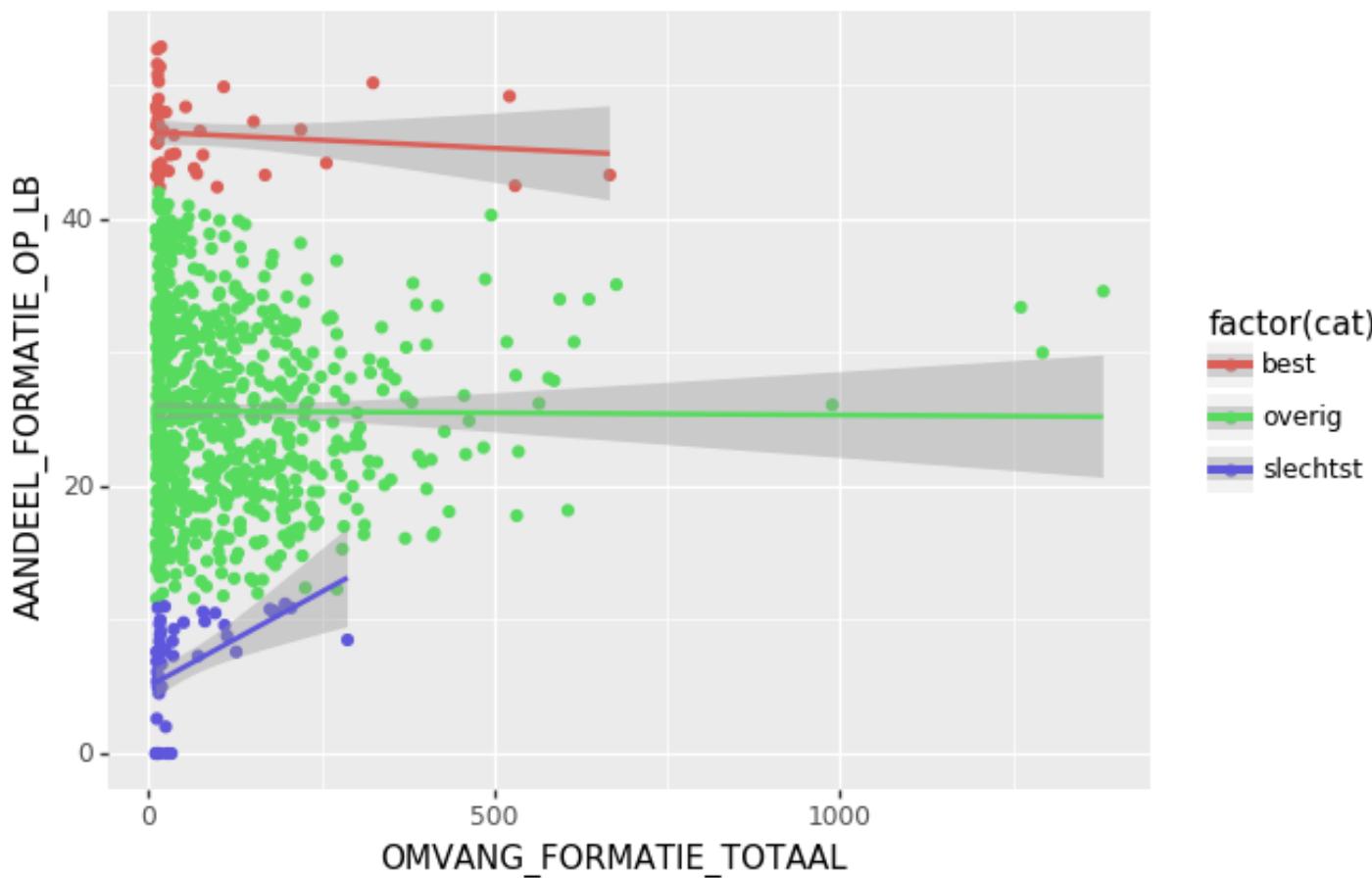
LOONSOM\_OP\_BRUTSAL (804 rows, -0.12/0.00/0.48 alles 0.03)



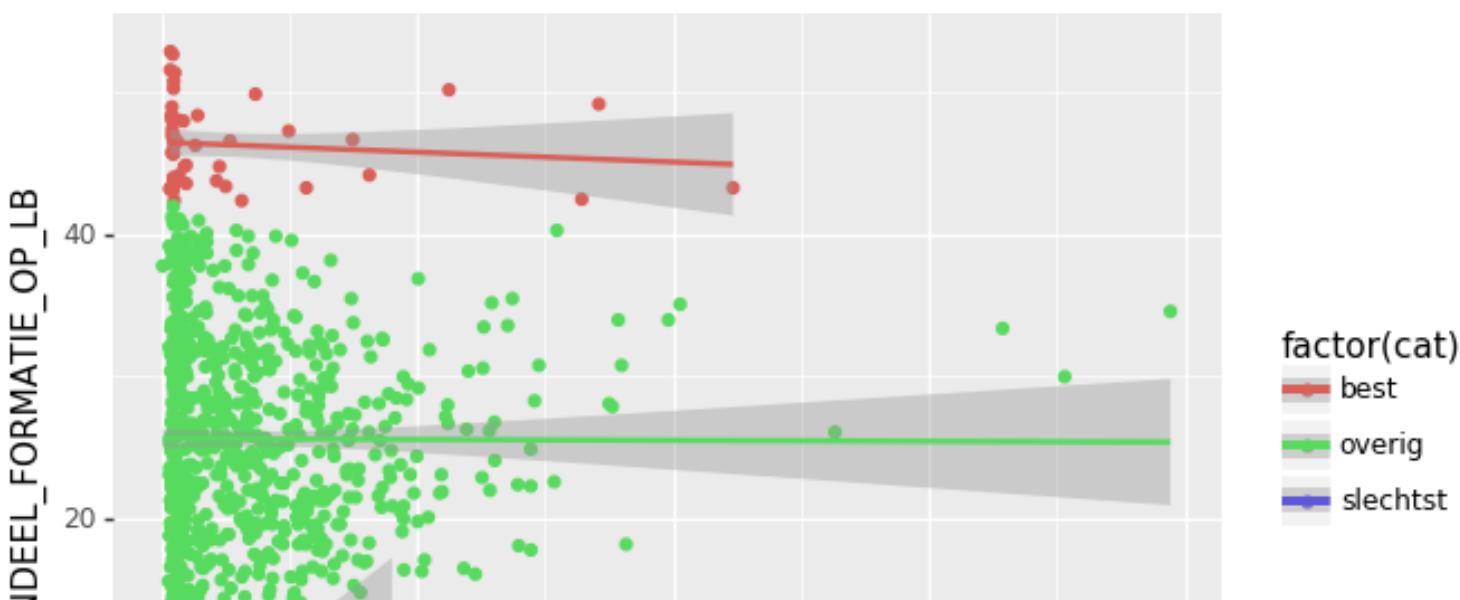
Aantal instellingen (801 rows, -0.06/-0.05/0.48 alles -0.01)



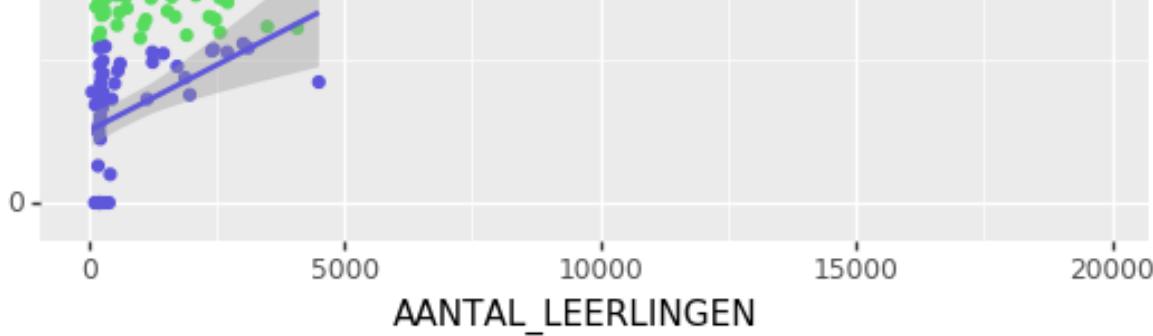
OMVANG\_FORMATIE\_TOTAAL (804 rows, -0.12/-0.01/0.48 alles 0.02)



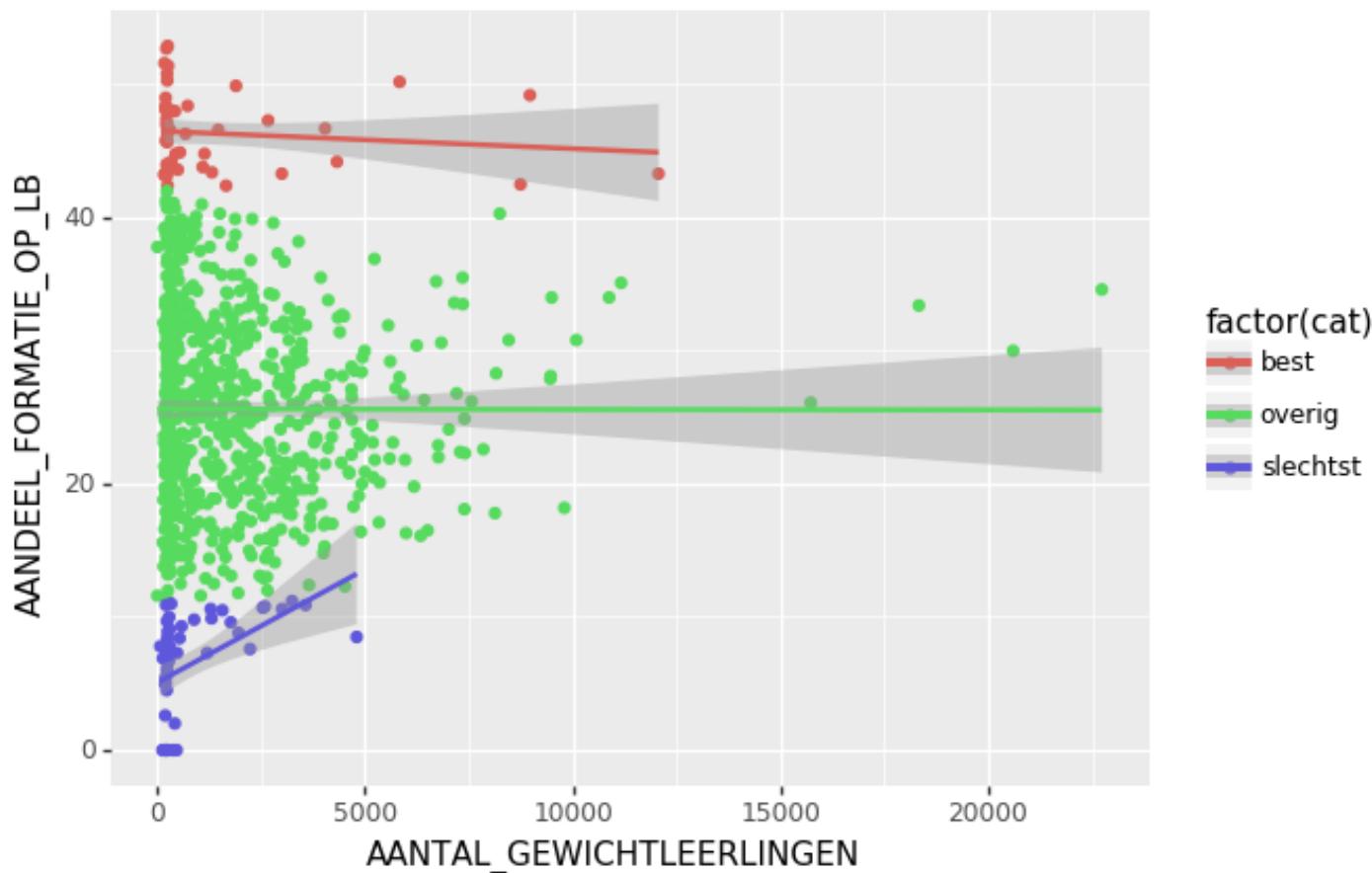
AANTAL\_LEERLINGEN (804 rows, -0.11/-0.00/0.47 alles 0.03)



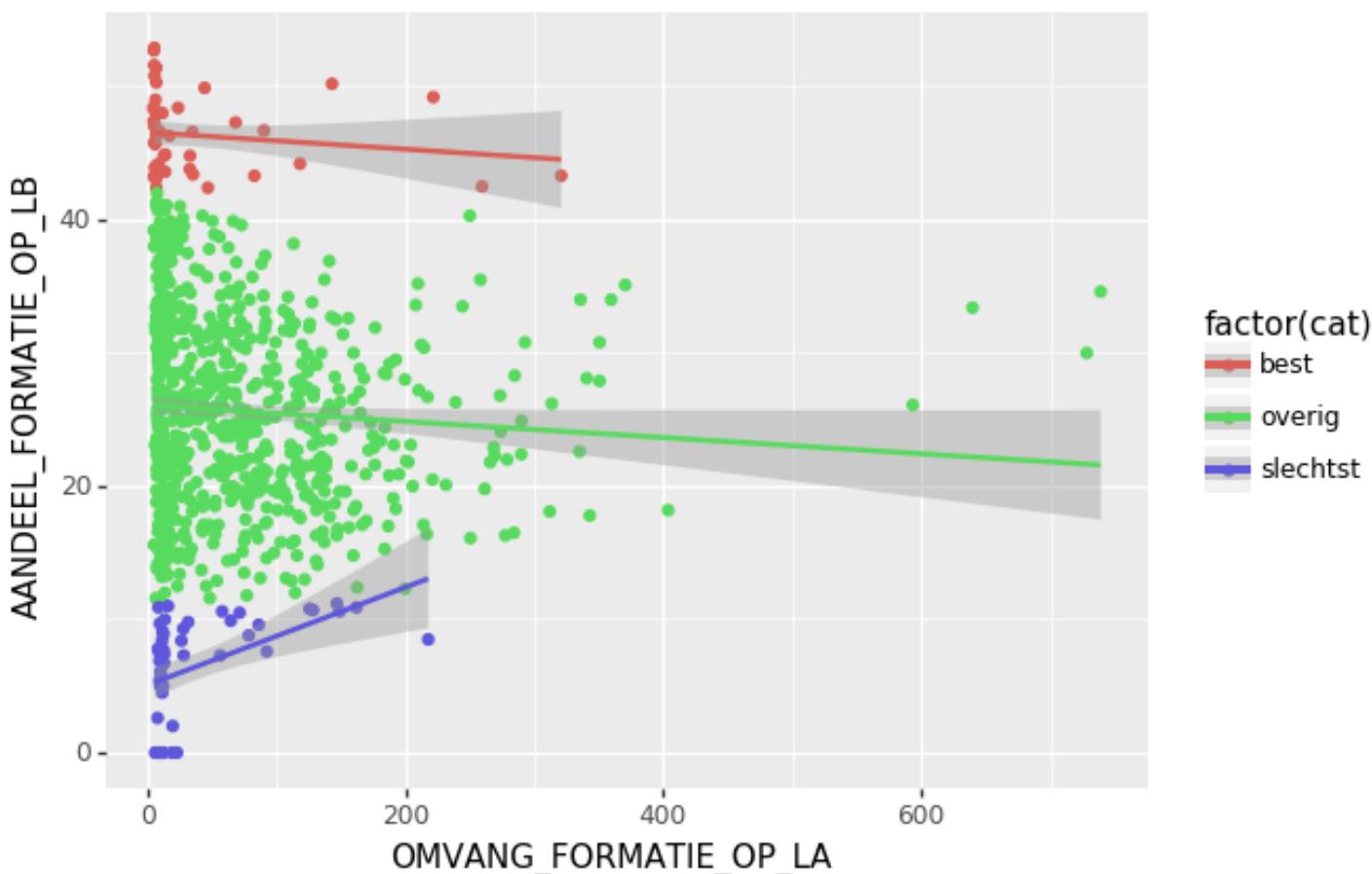
AAN



AANTAL\_GEWICHTLEERLINGEN (804 rows, -0.12/-0.00/0.47 alles 0.03)

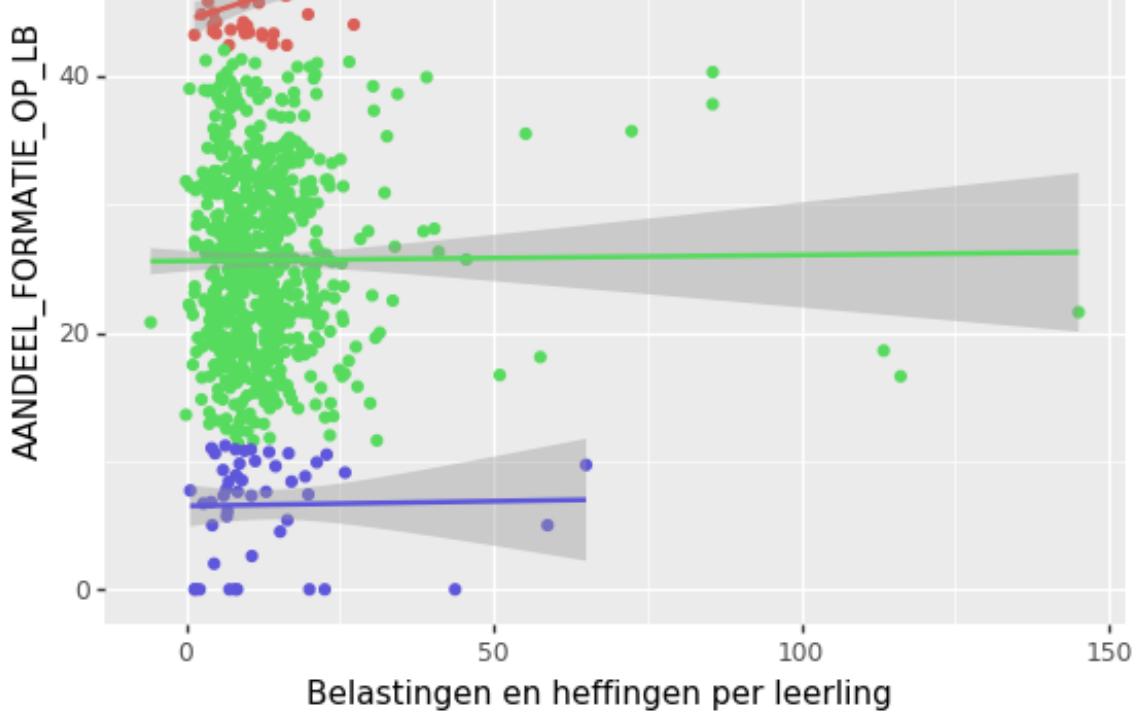


OMVANG\_FORMATIE\_OP\_LA (804 rows, -0.15/-0.07/0.47 alles -0.05)

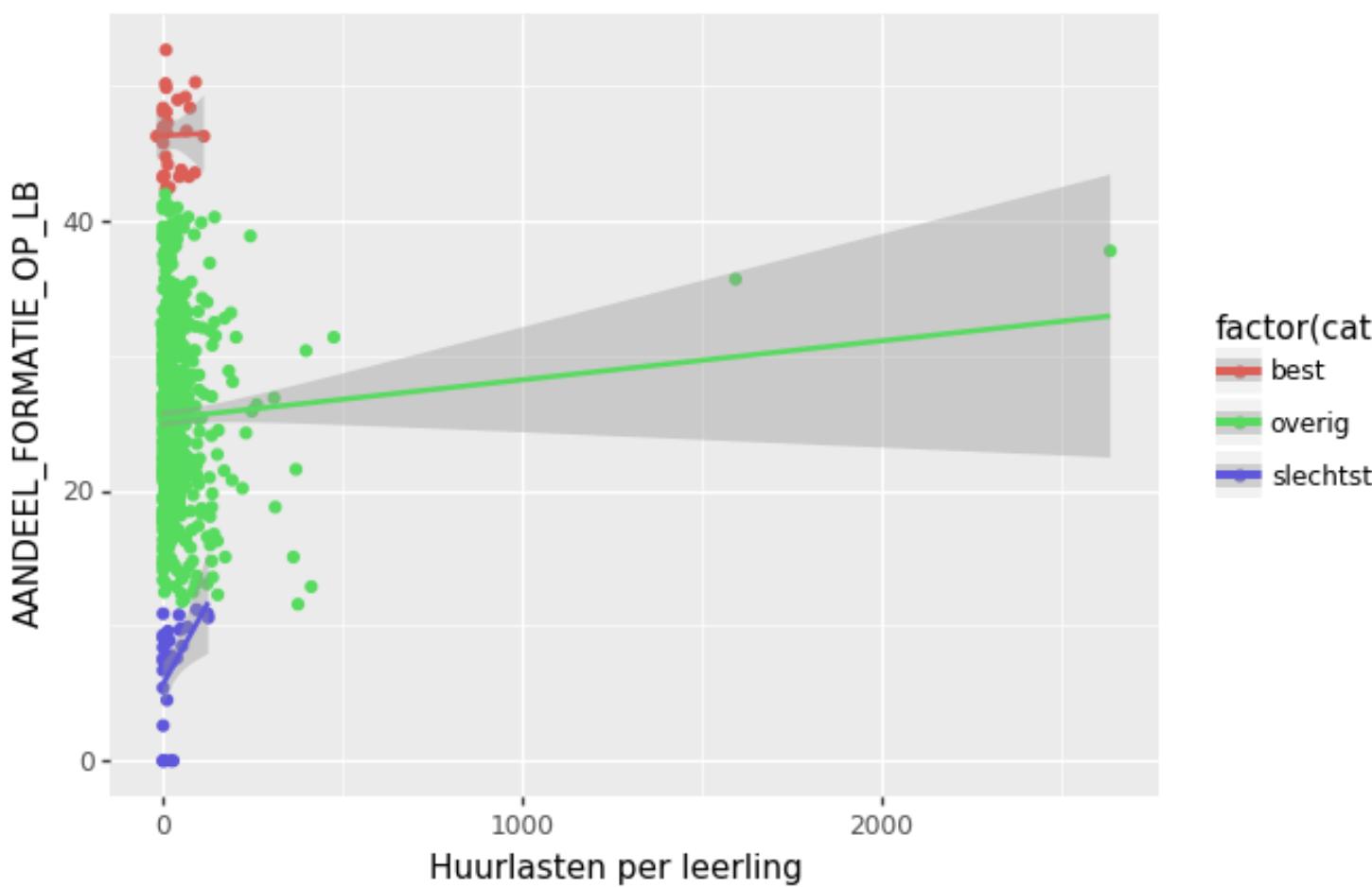


Belastingen en heffingen per leerling (779 rows, 0.46/0.01/0.03 alles 0.01)

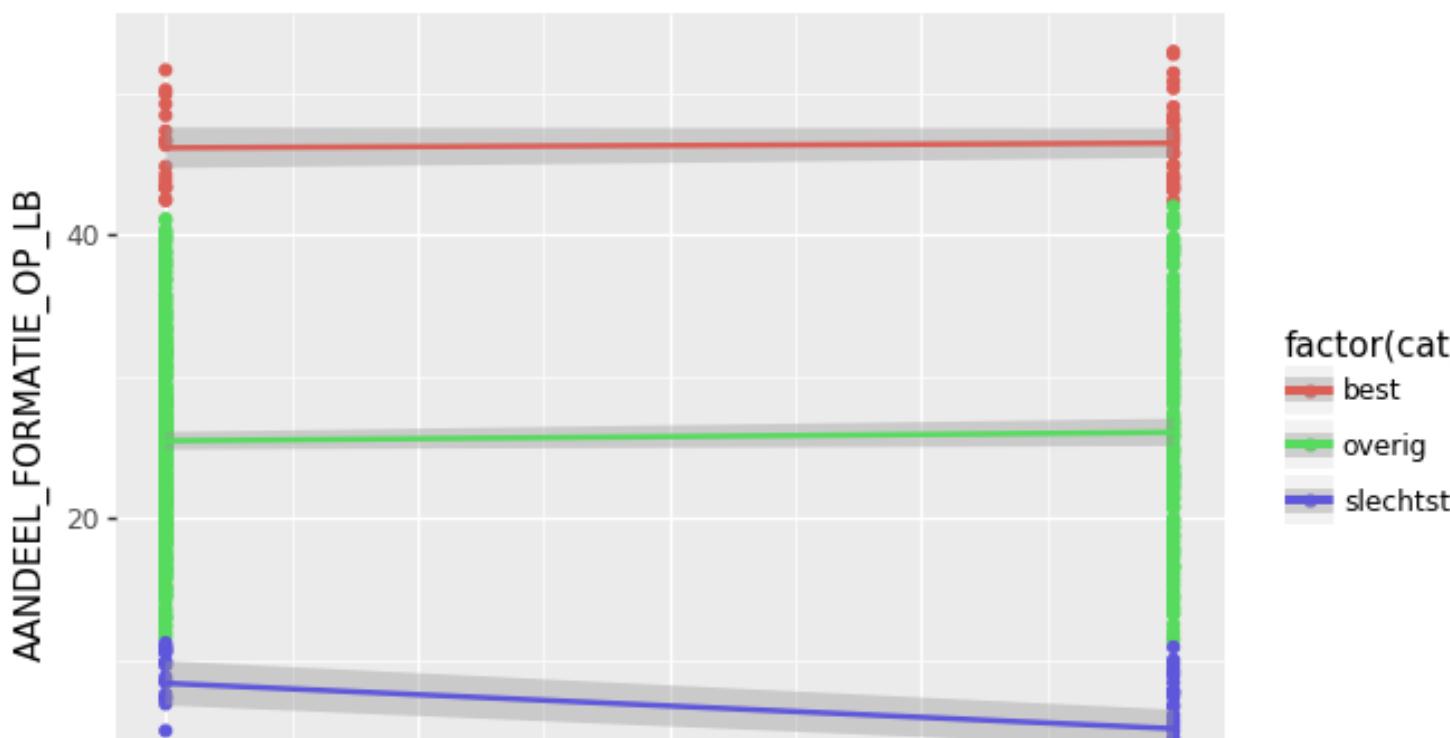


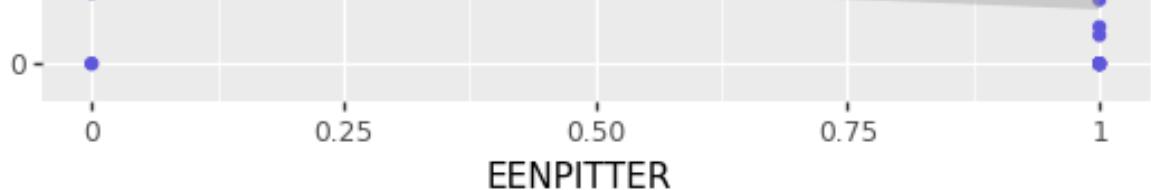


Huurlasten per leerling (600 rows, 0.02/0.06/0.46 alles 0.04)

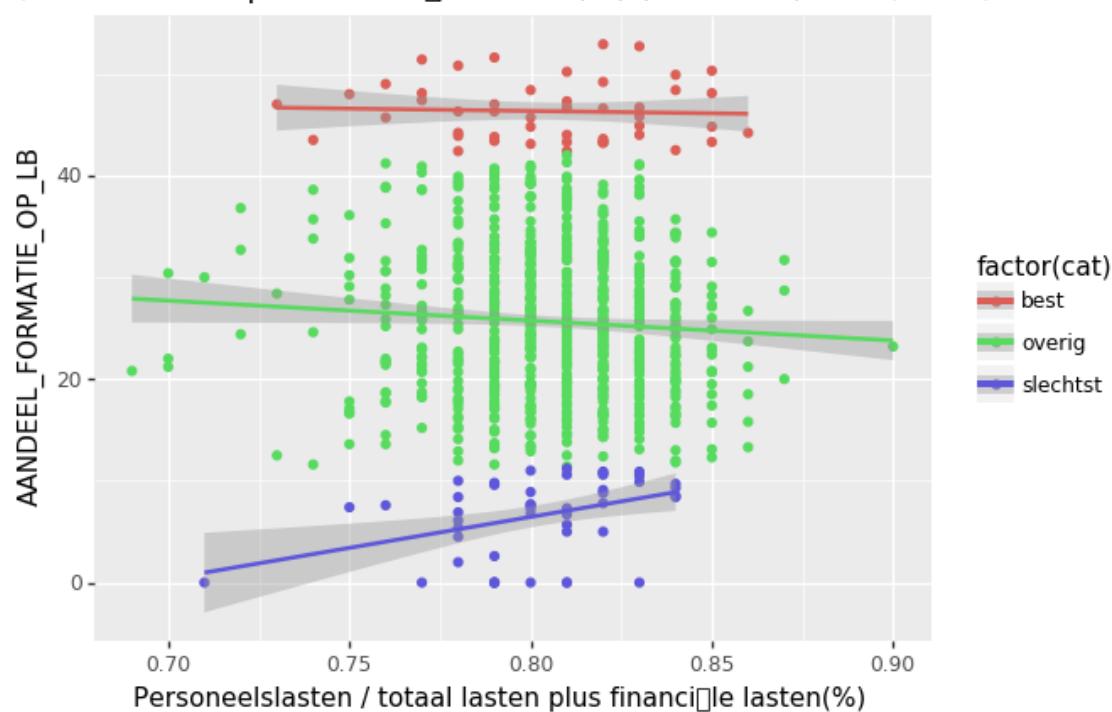


EENPITTER (804 rows, 0.05/0.04/-0.41 alles 0.04)





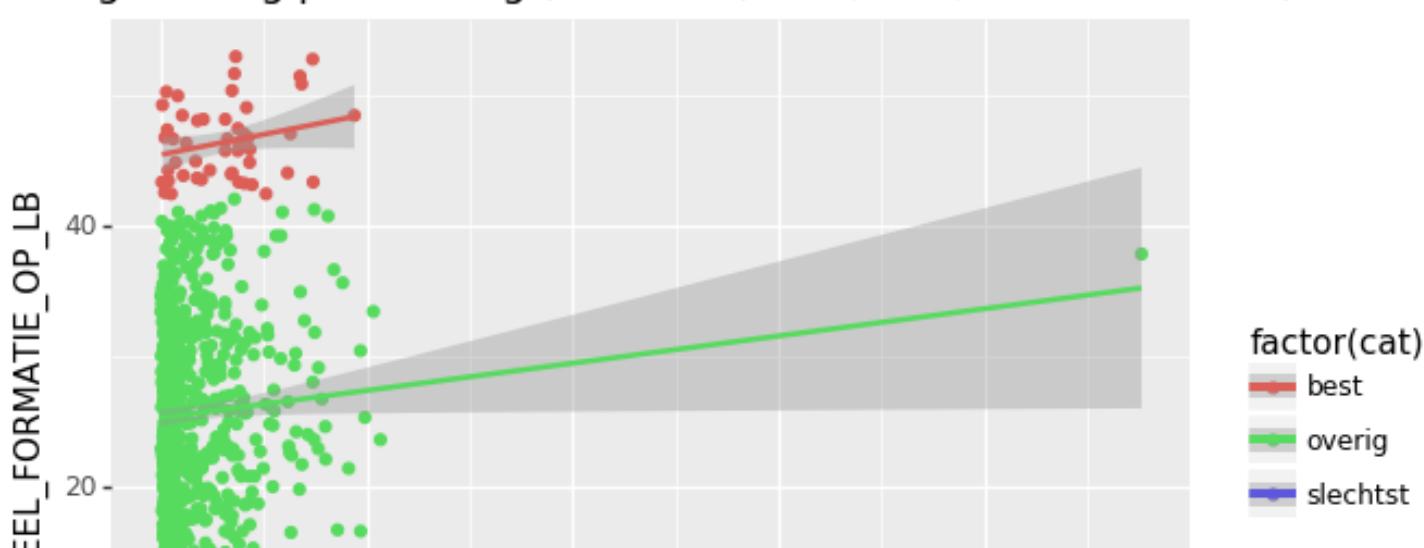
Personelelasten / totaal lasten plus financiële lasten(%) (791 rows, -0.05/-0.07/0.41 alles -0.03)

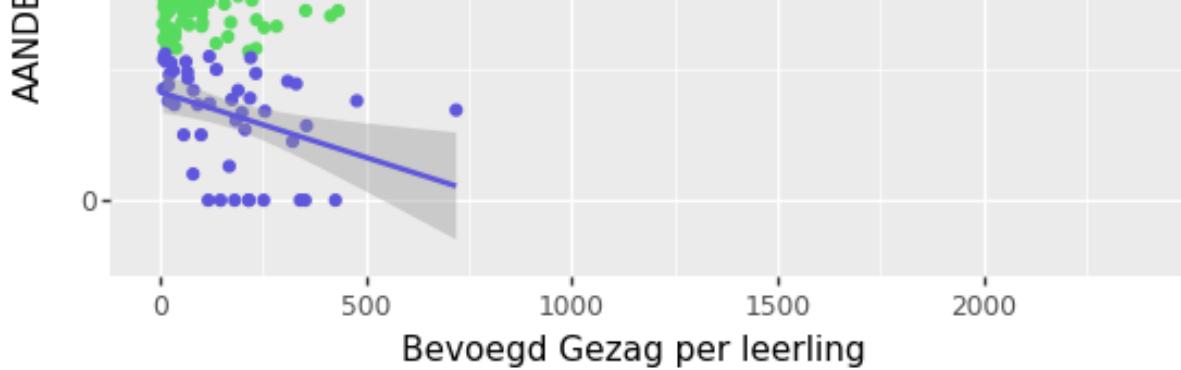


GEMIDDELDE\_LOONSOM\_OP\_BRUTSAL (804 rows, 0.26/0.25/0.40 alles 0.32)

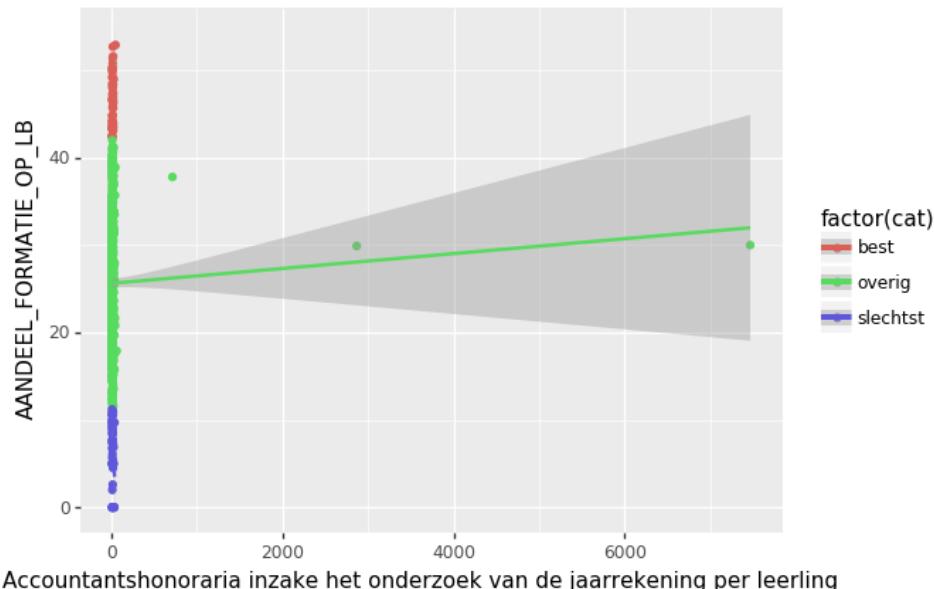


Bevoegd Gezag per leerling (791 rows, 0.25/0.08/-0.38 alles 0.03)

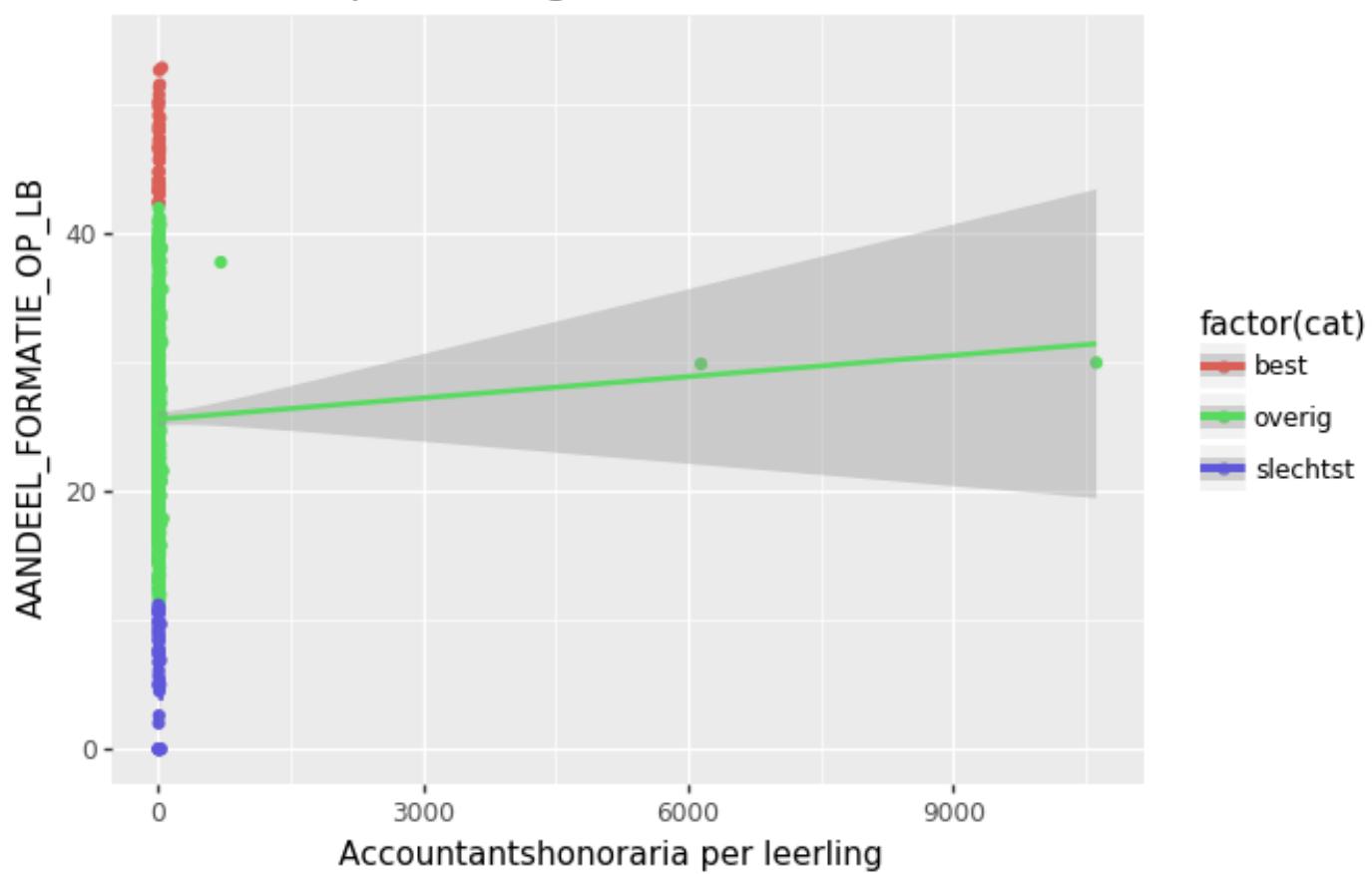




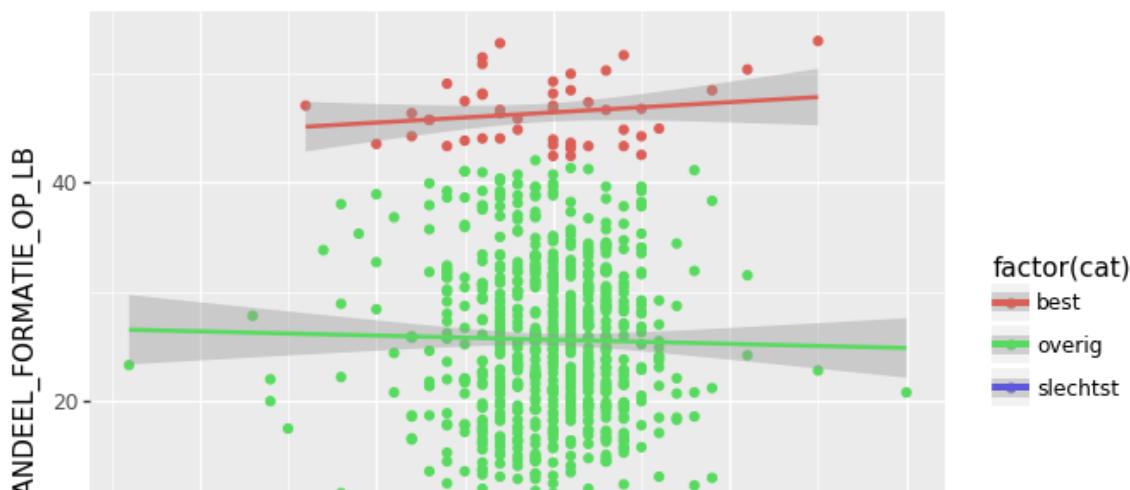
Accountantshonoraria inzake het onderzoek van de jaarrekening per leerling (784 rows, 0.31/0.04/-0.25 alles 0.02)

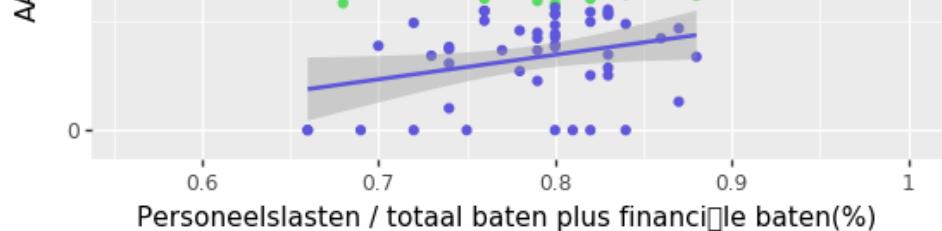


Accountantshonoraria per leerling (788 rows, 0.30/0.04/-0.24 alles 0.02)

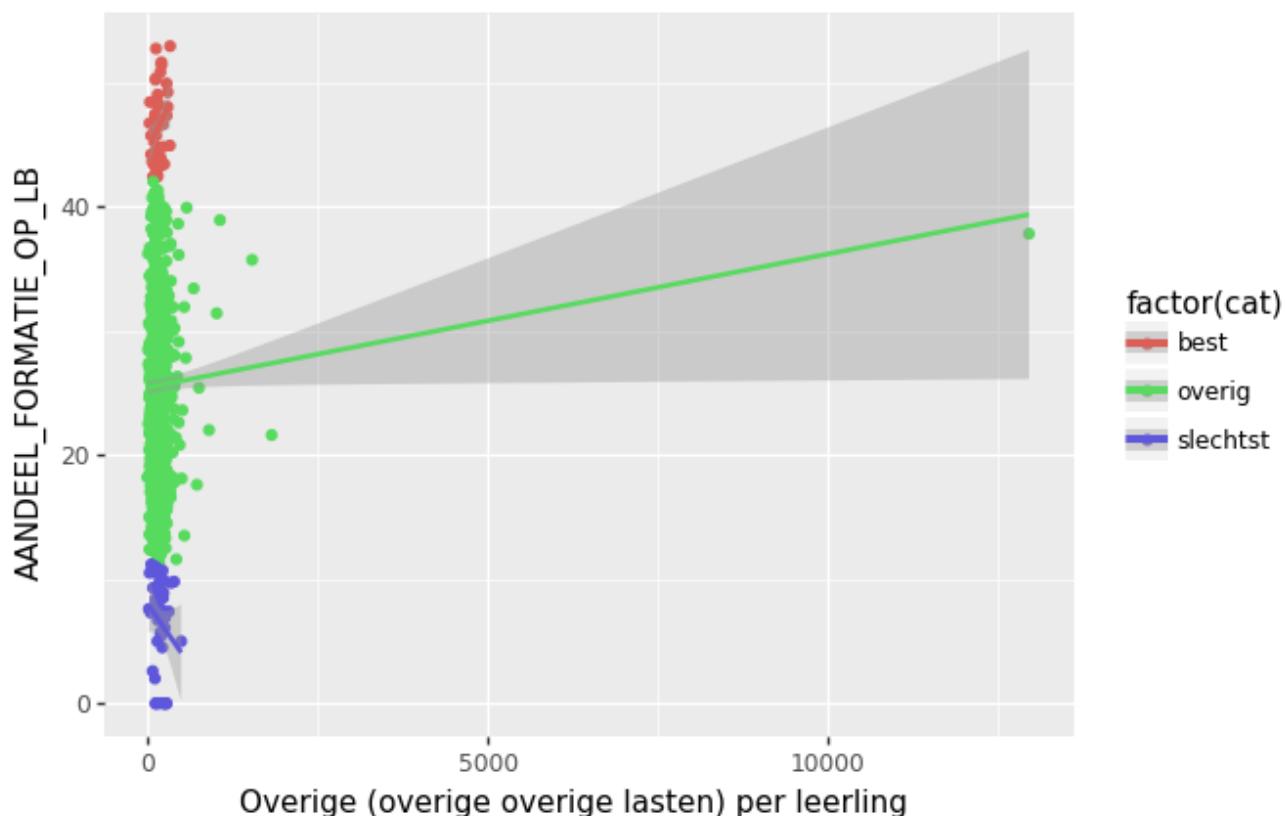


Personelelasten / totaal baten plus financiële baten(%) (791 rows, 0.17/-0.02/0.30 alles 0.02)

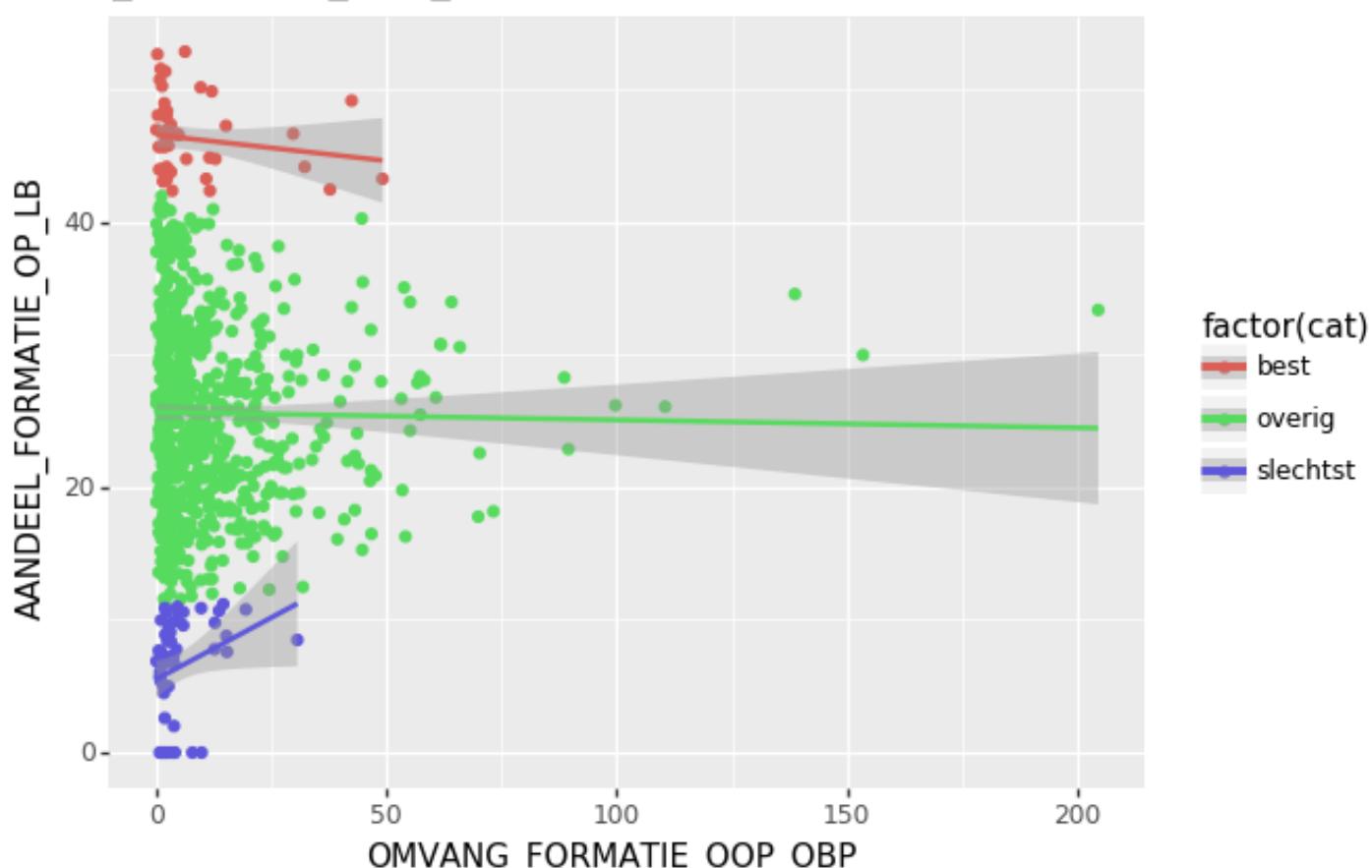




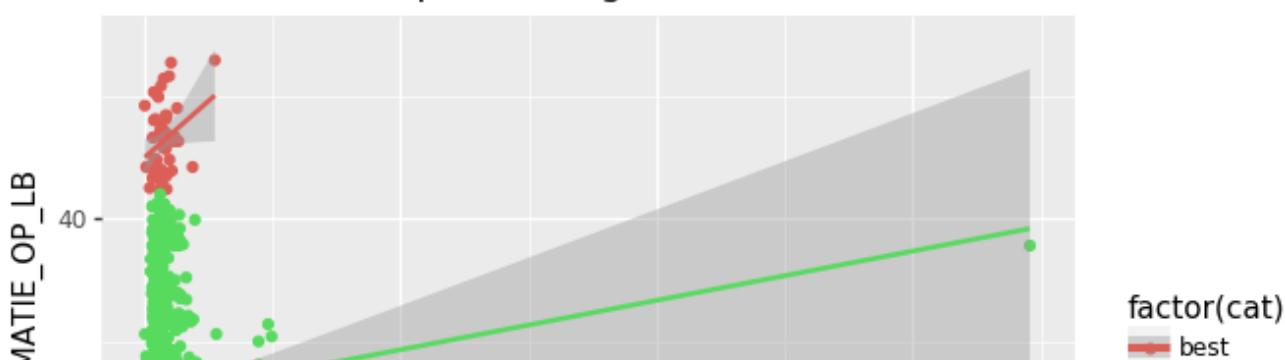
Overige (overige overige lasten) per leerling (774 rows, 0.29/0.08/-0.20 alles 0.05)

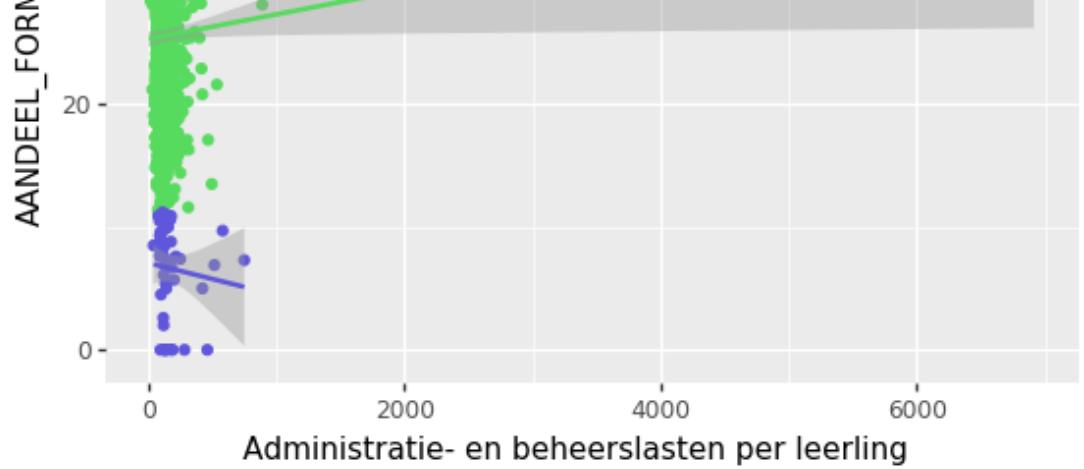


OMVANG\_FORMATIE\_OOP\_OBP (804 rows, -0.15/-0.01/0.29 alles 0.00)

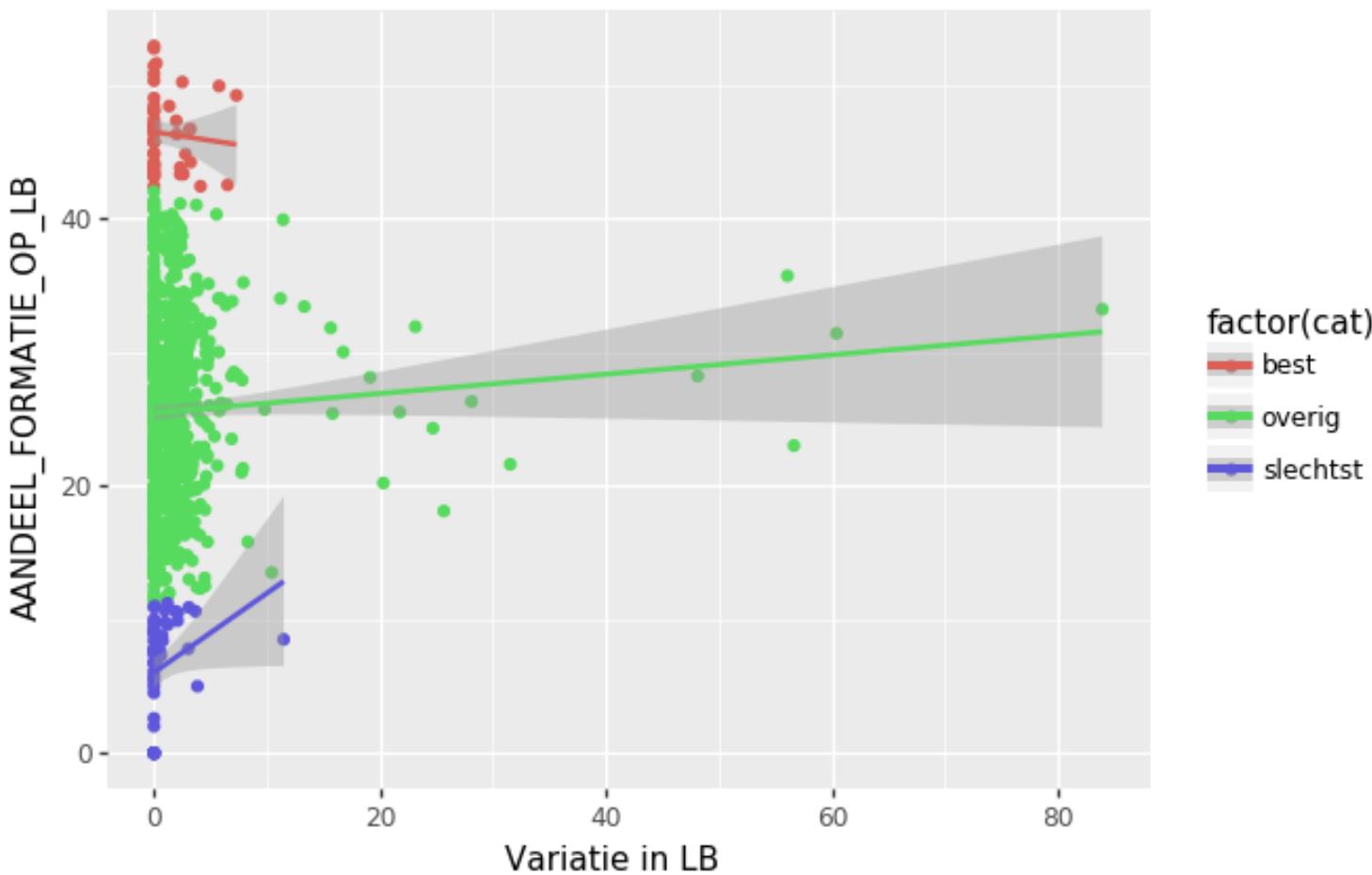


Administratie- en beheerslasten per leerling (791 rows, 0.29/0.08/-0.09 alles 0.04)

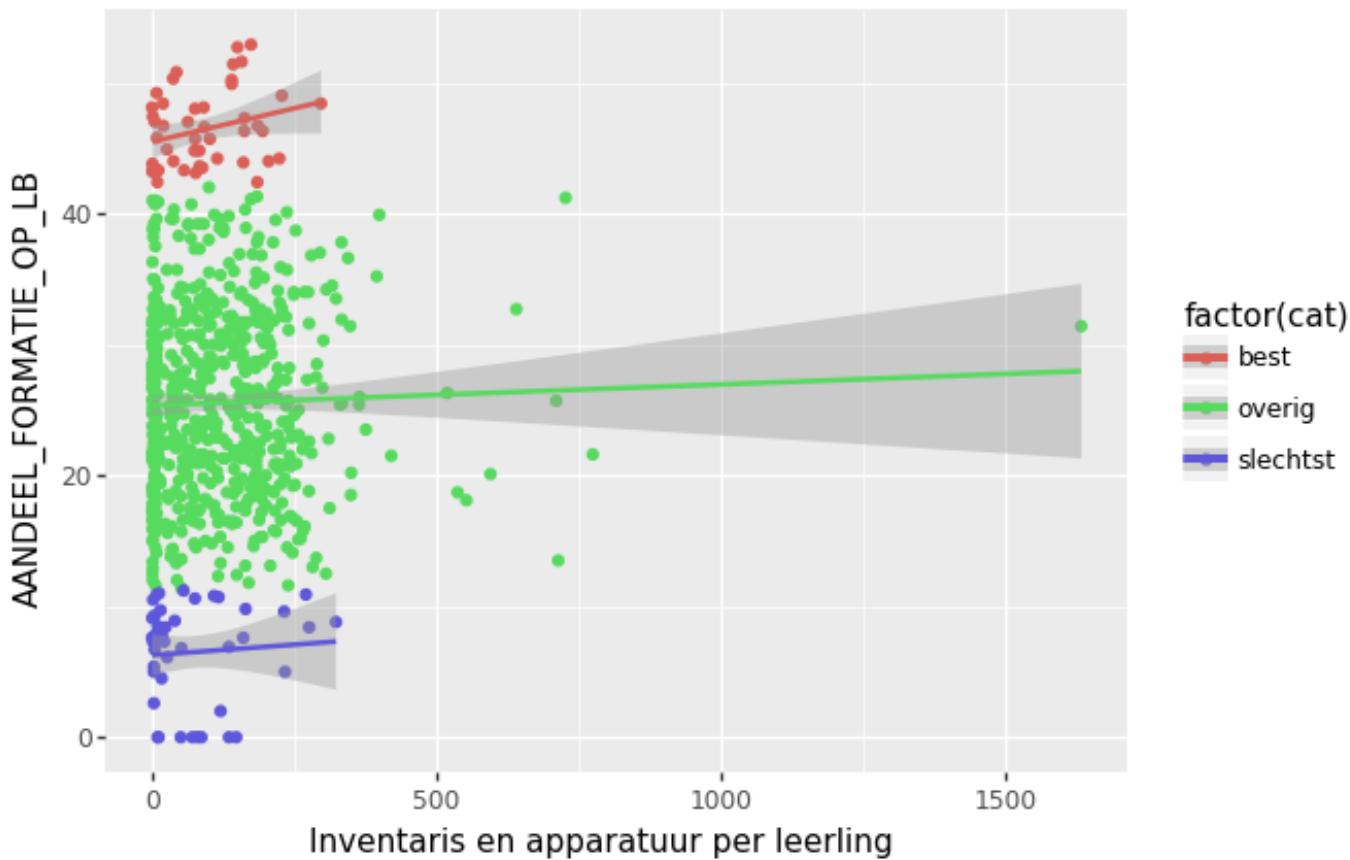




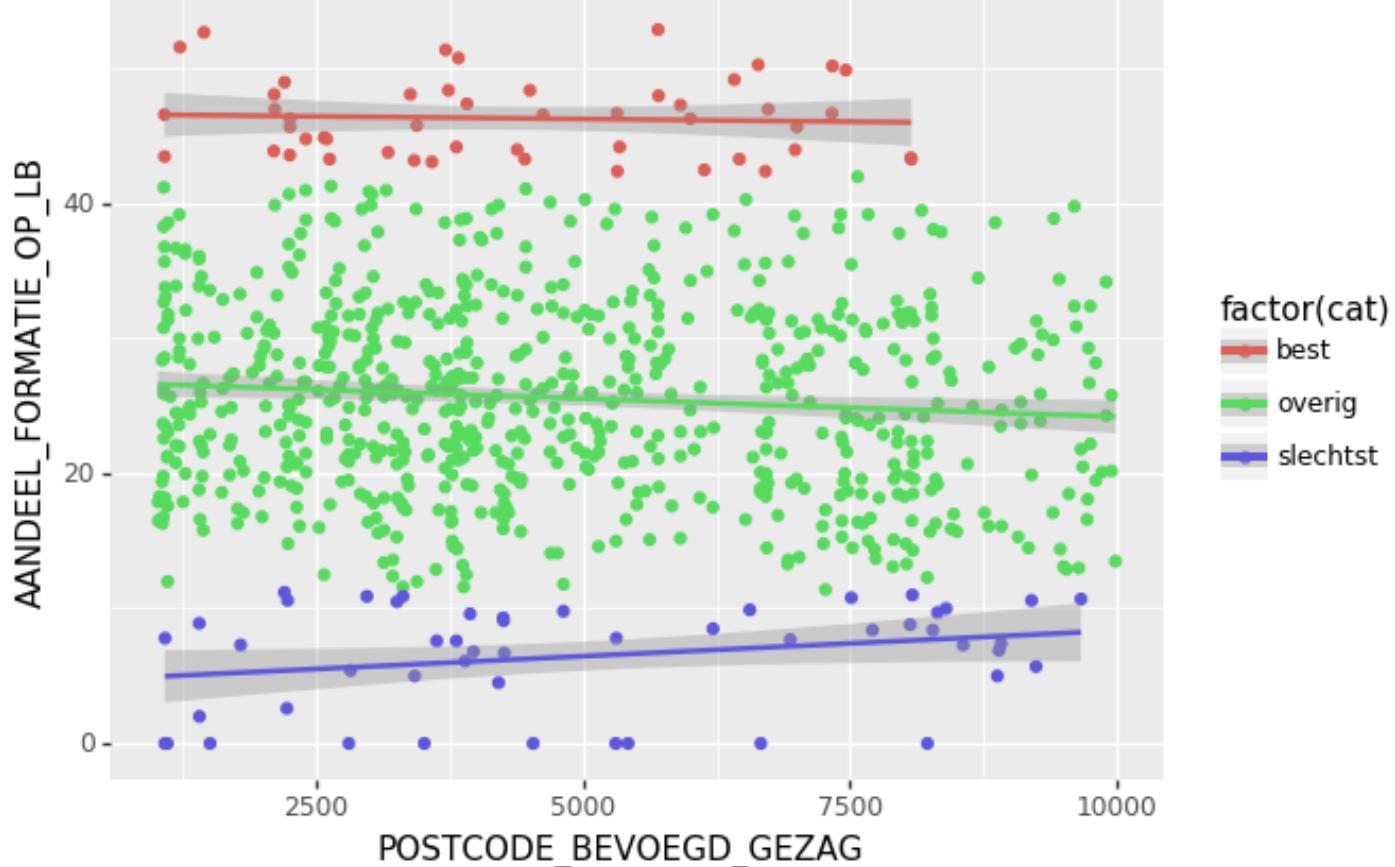
Variatie in LB (804 rows, -0.08/0.06/0.28 alles 0.05)



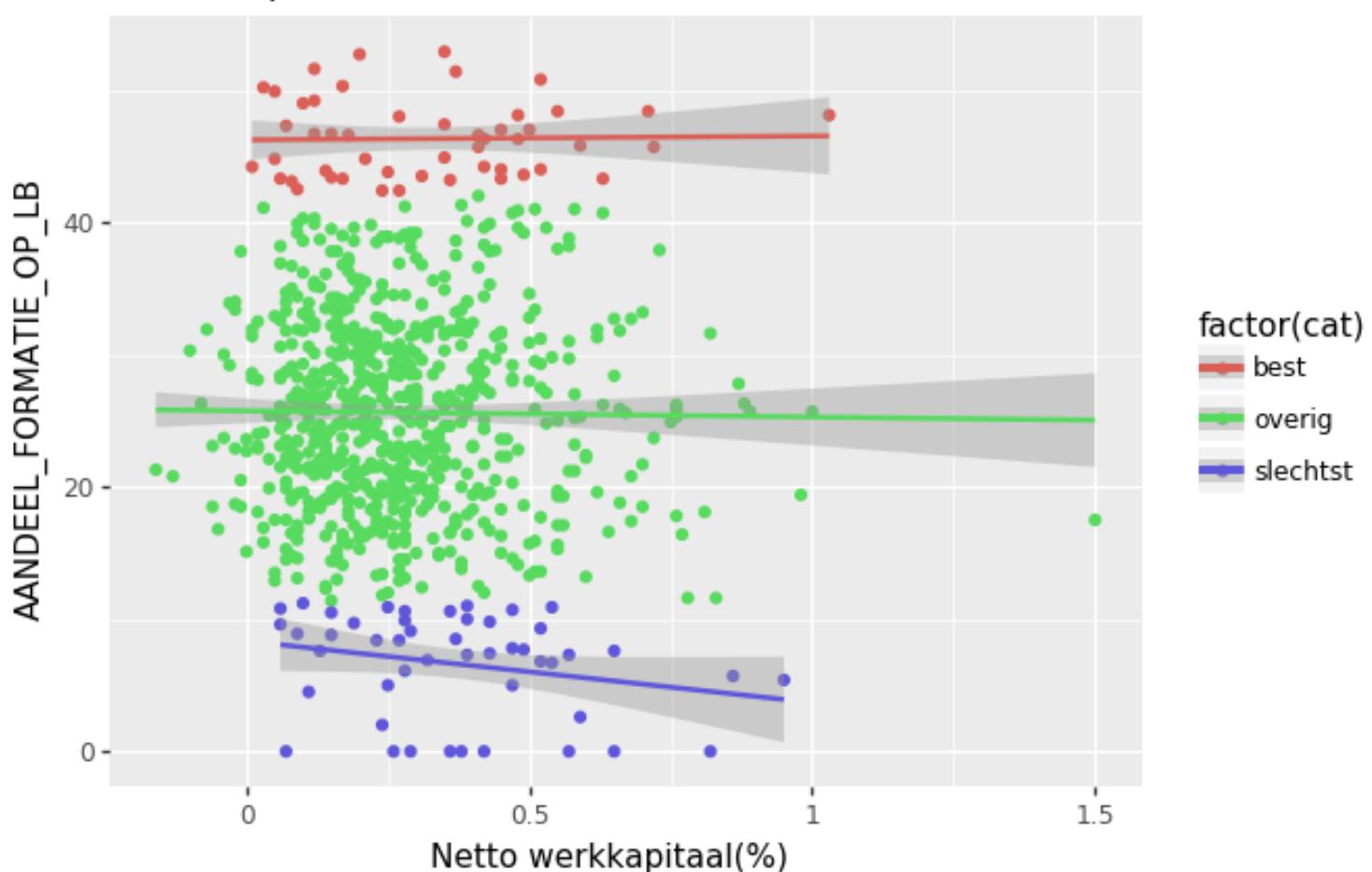
Inventaris en apparatuur per leerling (695 rows, 0.27/0.03/0.07 alles 0.03)



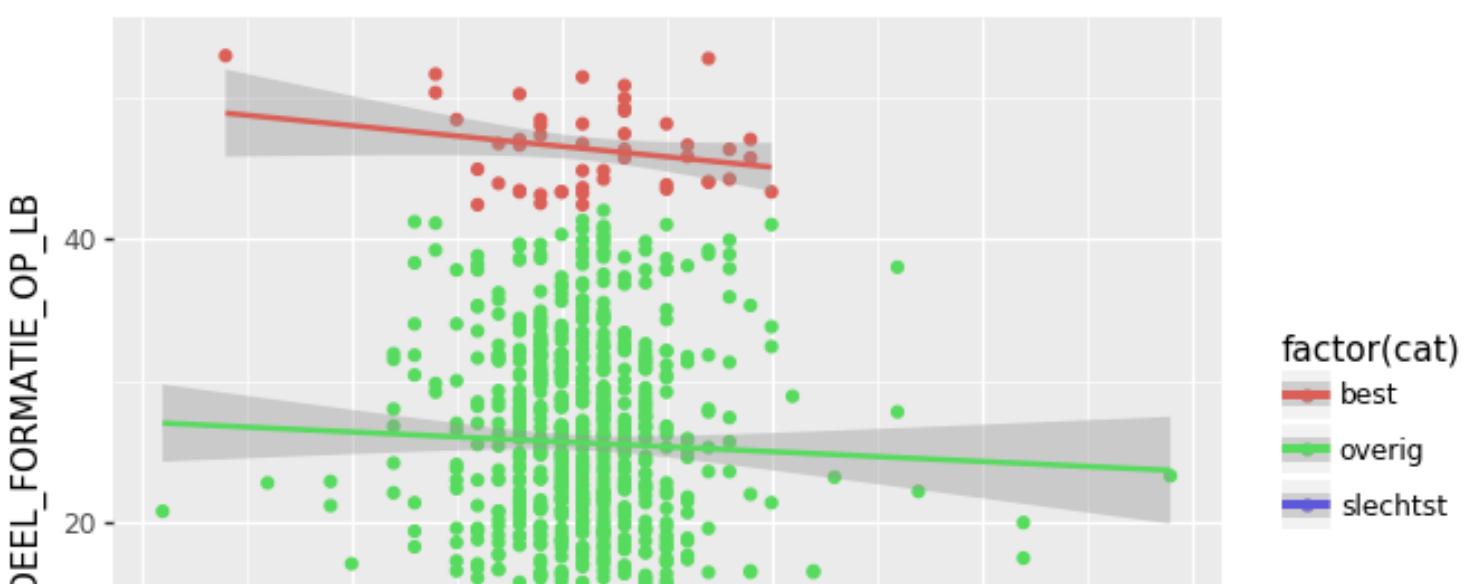
POSTCODE\_BEVOEGD\_GEZAG (804 rows, -0.06/-0.09/0.26 alles -0.09)

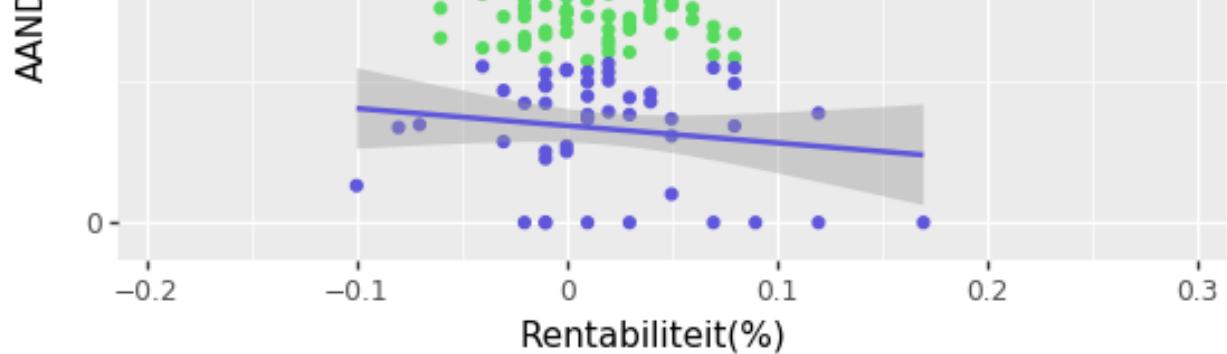


Netto werkkapitaal(%) (791 rows, 0.02/-0.01/-0.25 alles -0.05)

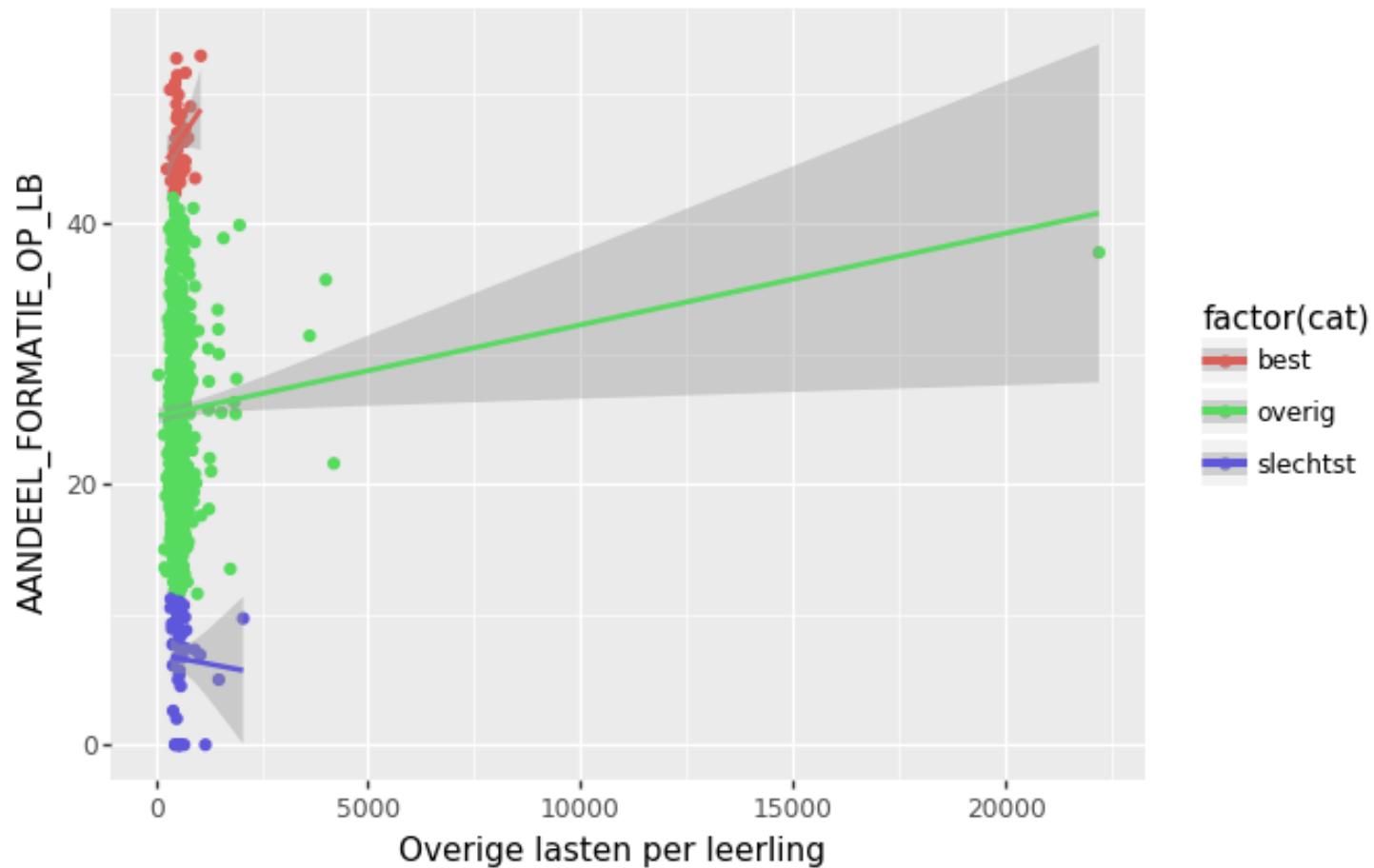


Rentabiliteit(%) (791 rows, -0.24/-0.04/-0.16 alles -0.05)

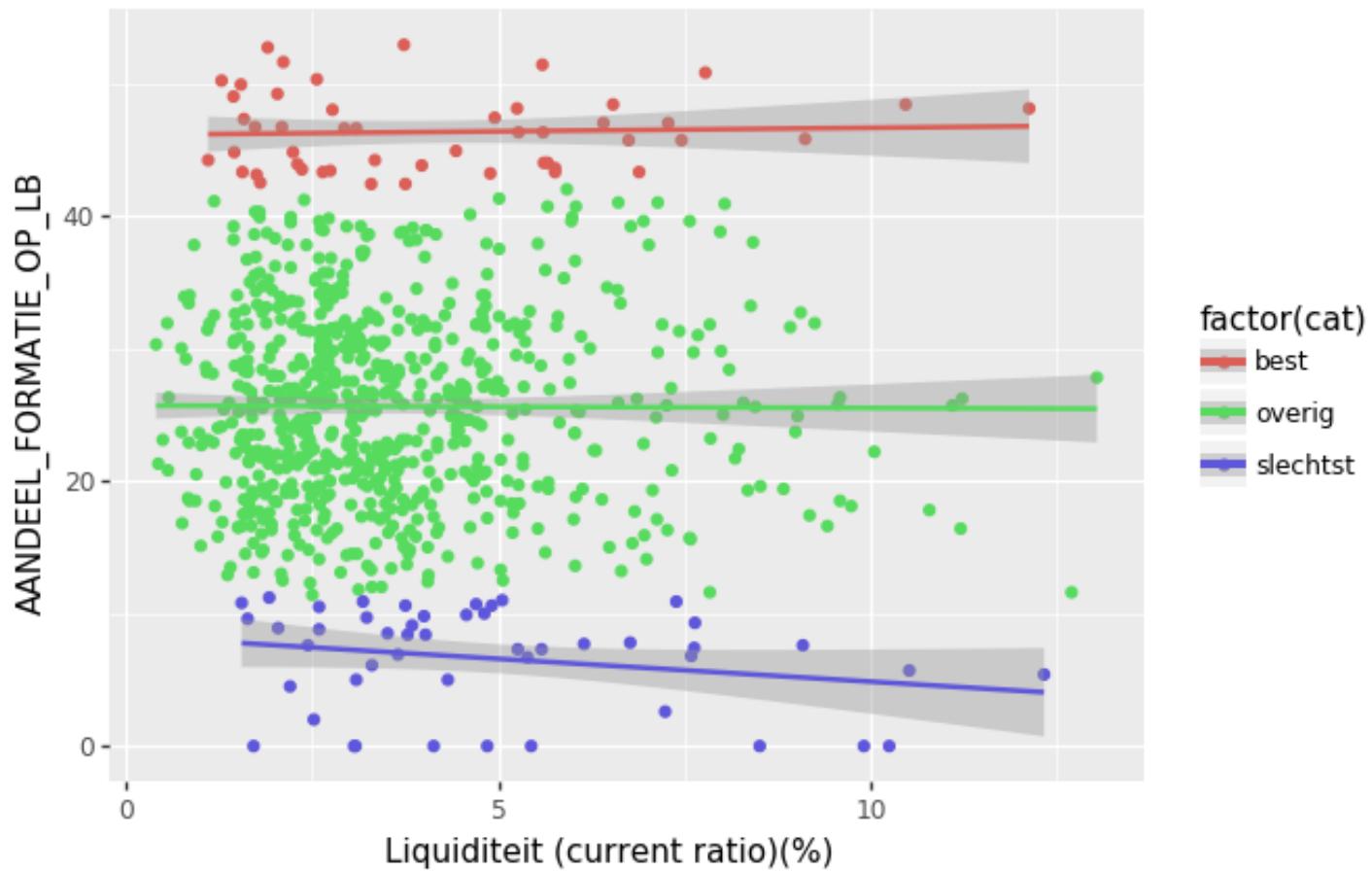




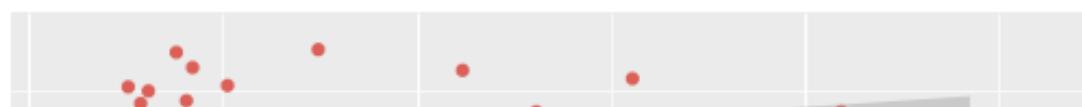
Overige lasten per leerling (791 rows, 0.23/0.09/-0.05 alles 0.05)

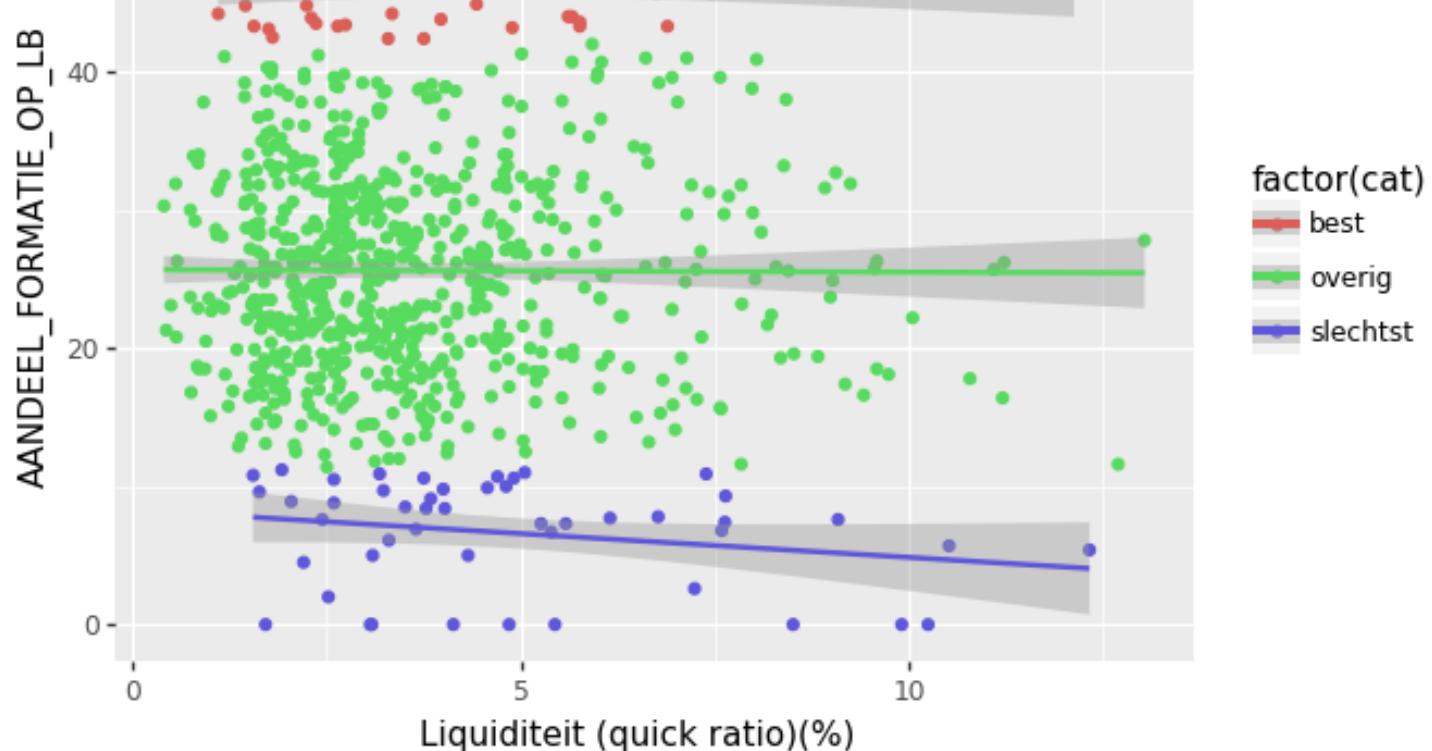


Liquiditeit (current ratio)(%) (791 rows, 0.05/-0.01/-0.23 alles -0.05)

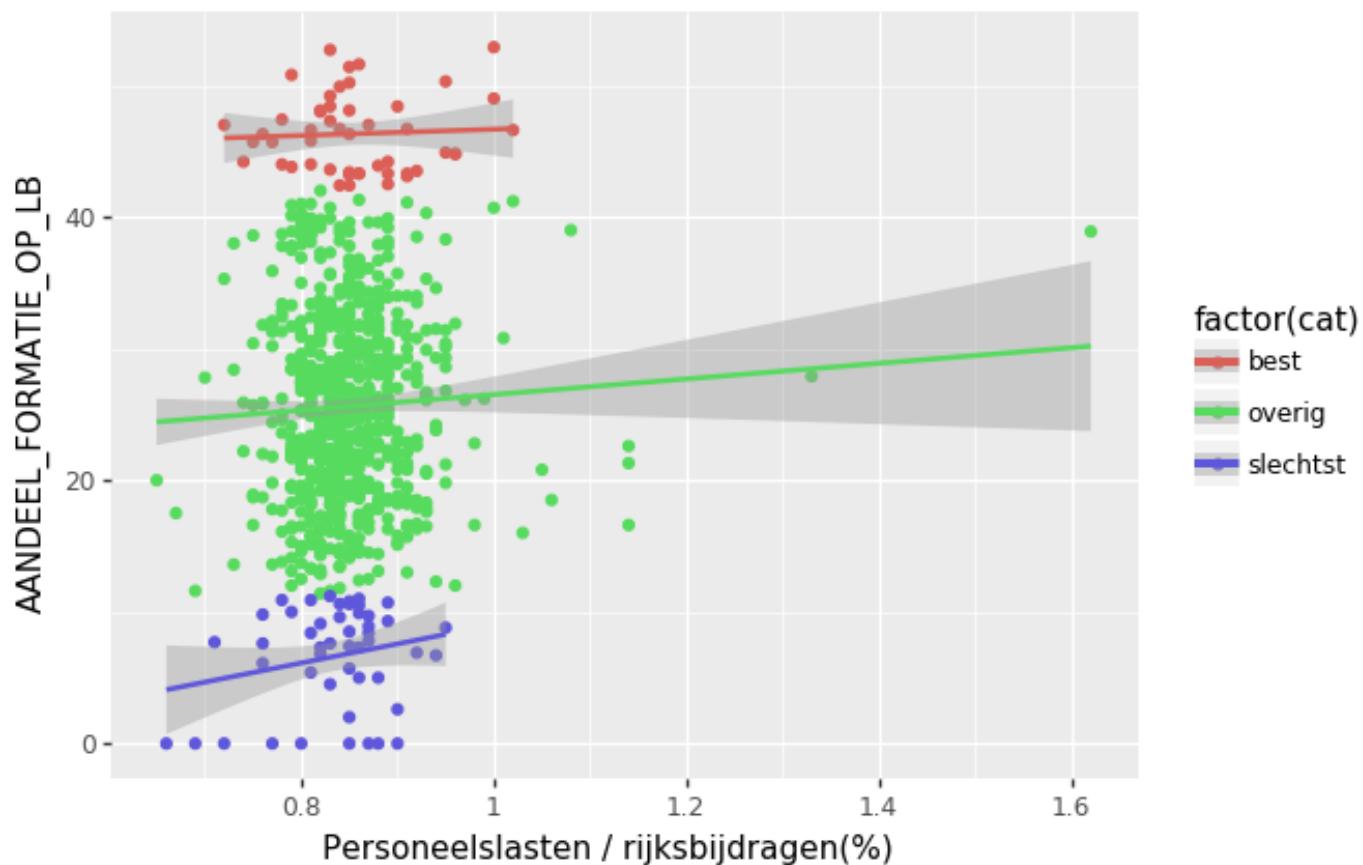


Liquiditeit (quick ratio)(%) (791 rows, 0.05/-0.01/-0.23 alles -0.05)

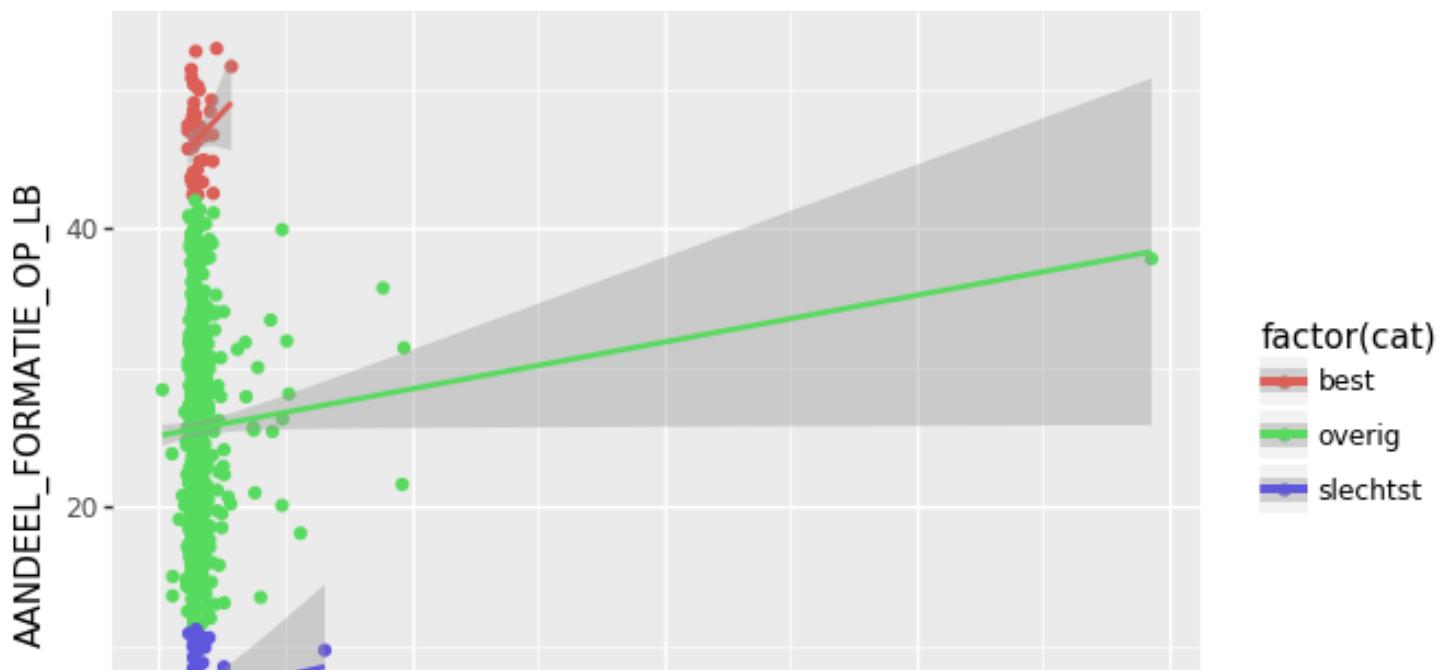




Personelelasten / riksbijdragen(%) (791 rows, 0.06/0.05/0.23 alles 0.08)

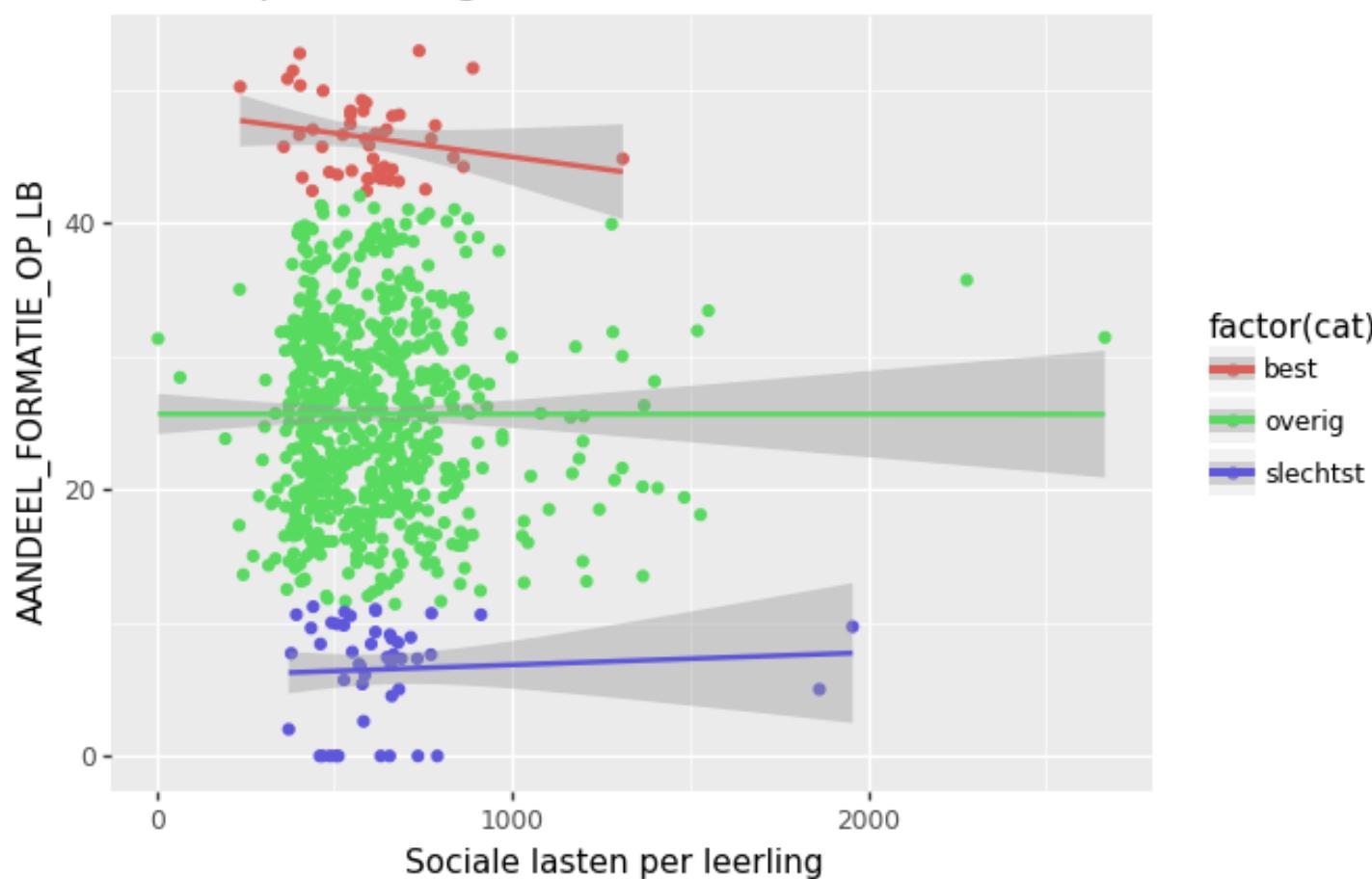


Pensioenlasten per leerling (767 rows, 0.23/0.08/0.10 alles 0.04)

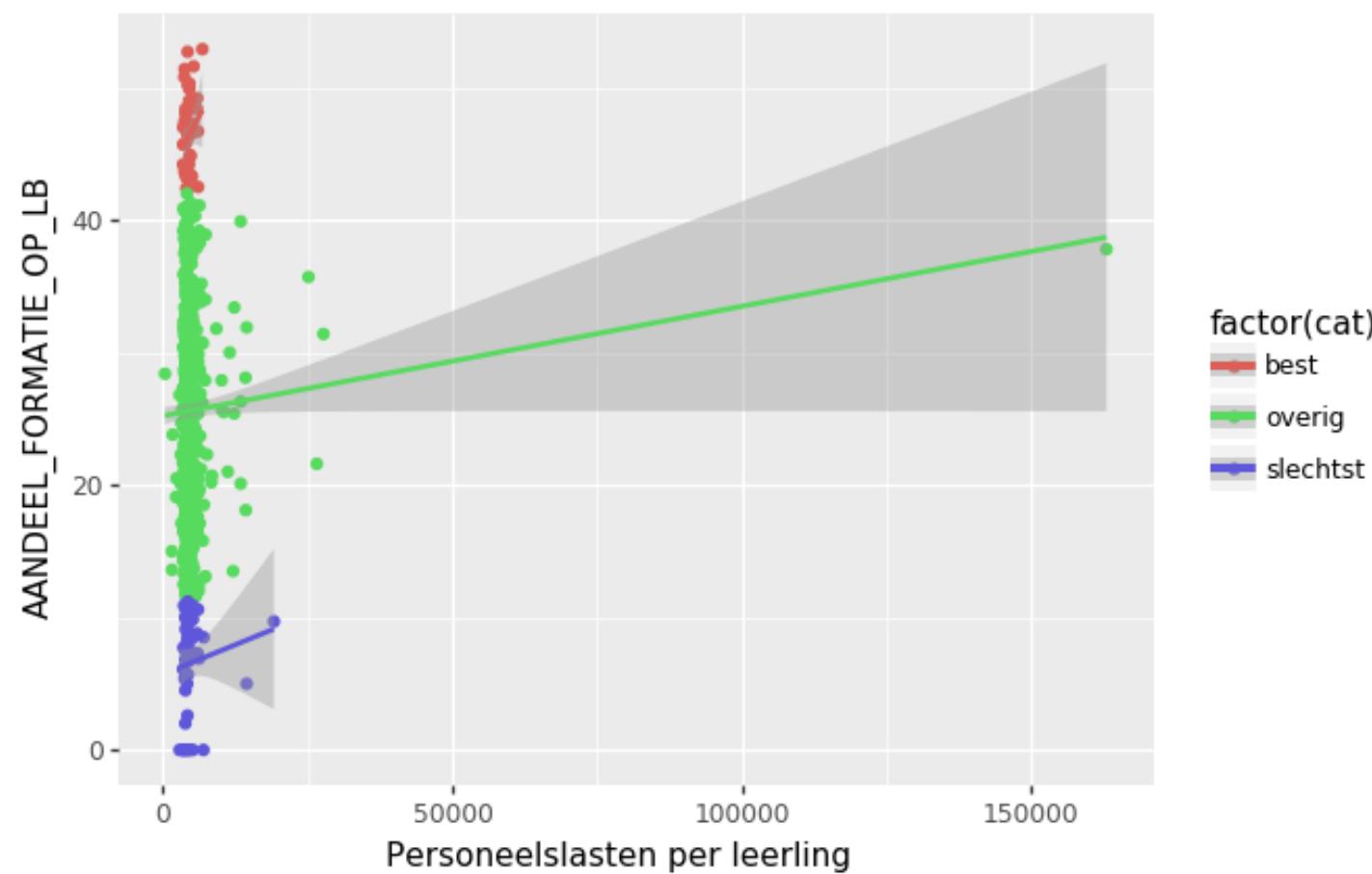




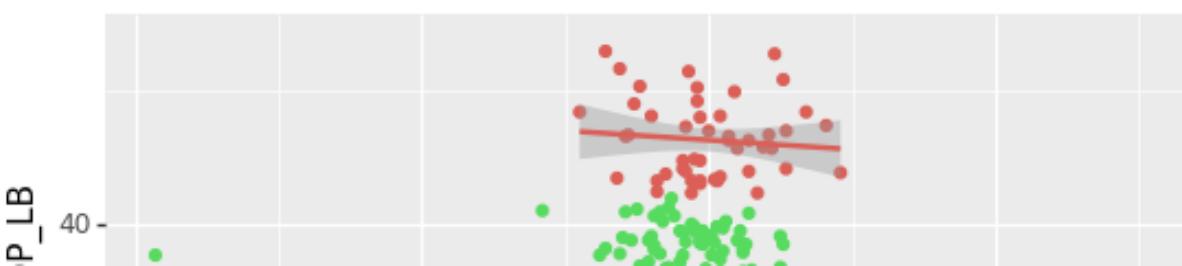
Sociale lasten per leerling (774 rows, -0.21/-0.00/0.07 alles -0.03)

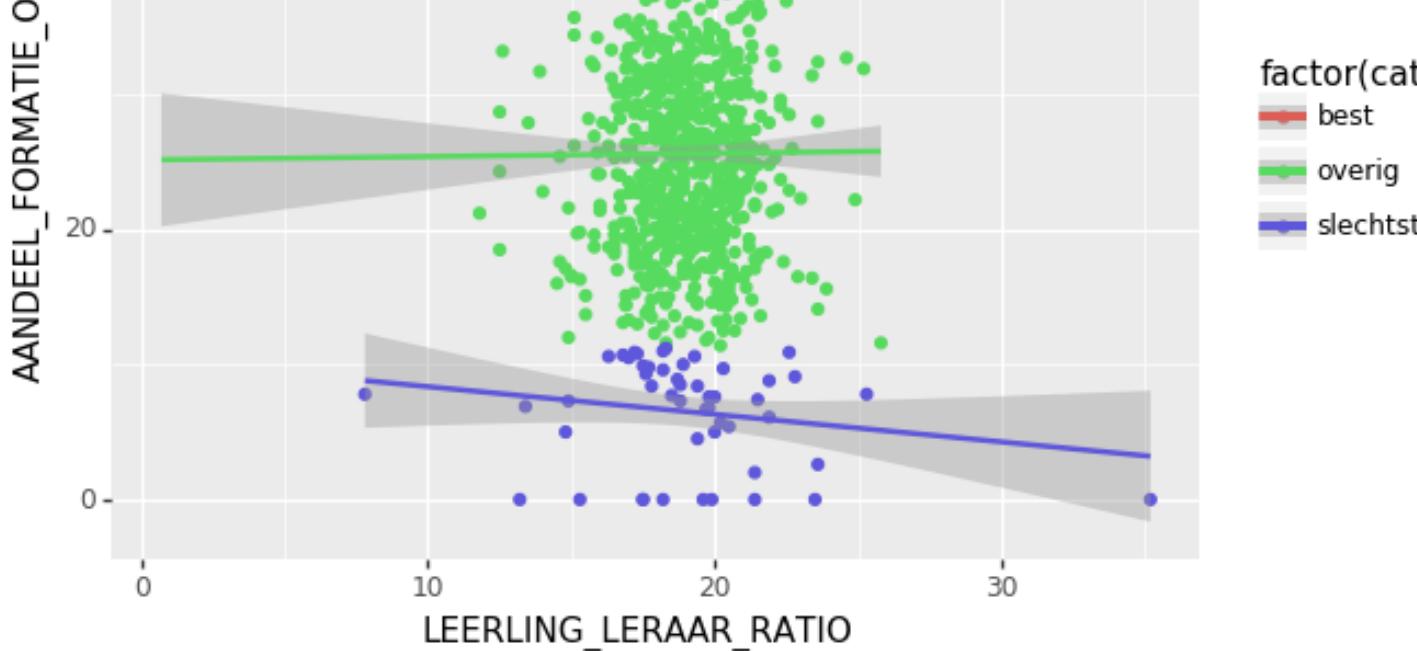


Personelelasten per leerling (791 rows, 0.21/0.07/0.13 alles 0.04)

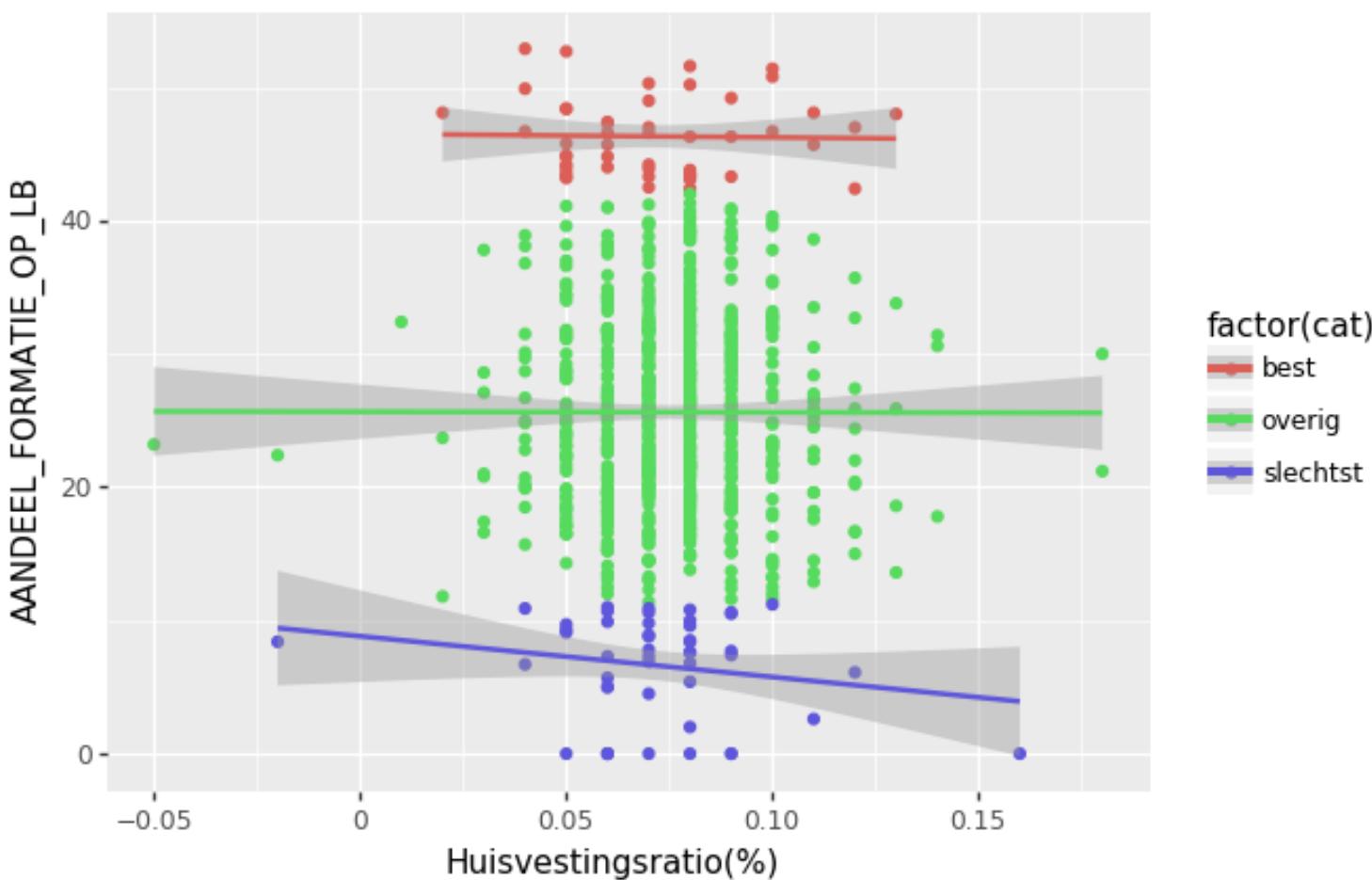


LEERLING\_LERAAR\_RATIO (804 rows, -0.10/0.01/-0.20 alles 0.05)

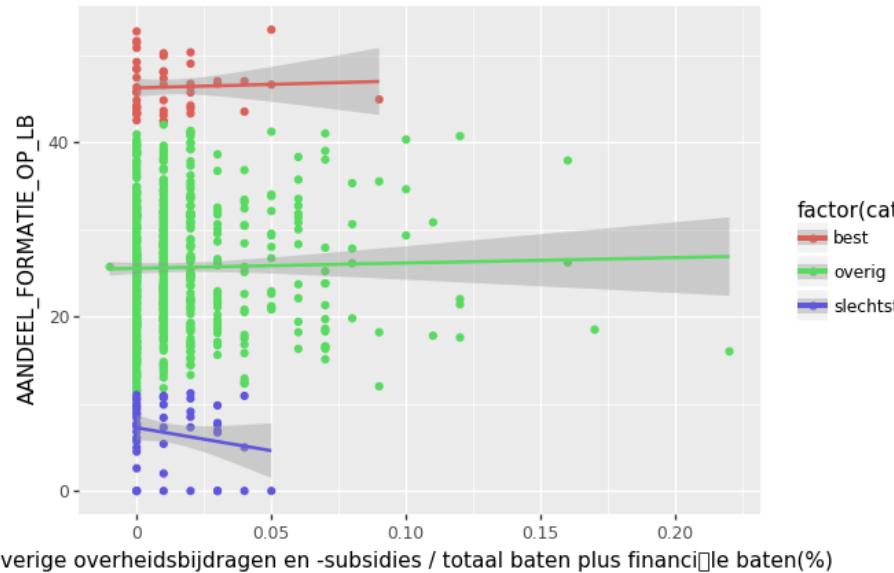




Huisvestingsratio(%) (791 rows, -0.02/-0.00/-0.20 alles -0.02)

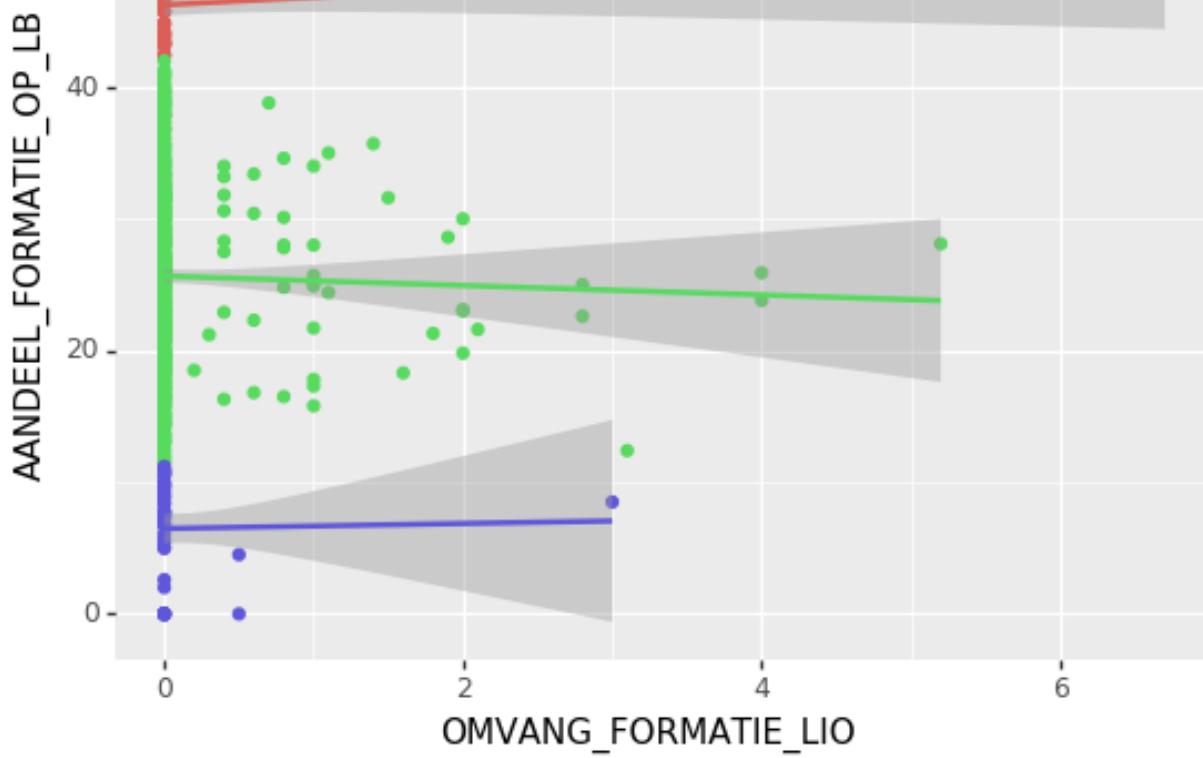


Overige overheidsbijdragen en -subsidies / totaal baten plus financiële baten(%) (791 rows, 0.05/0.02/-0.20 alles 0.01)

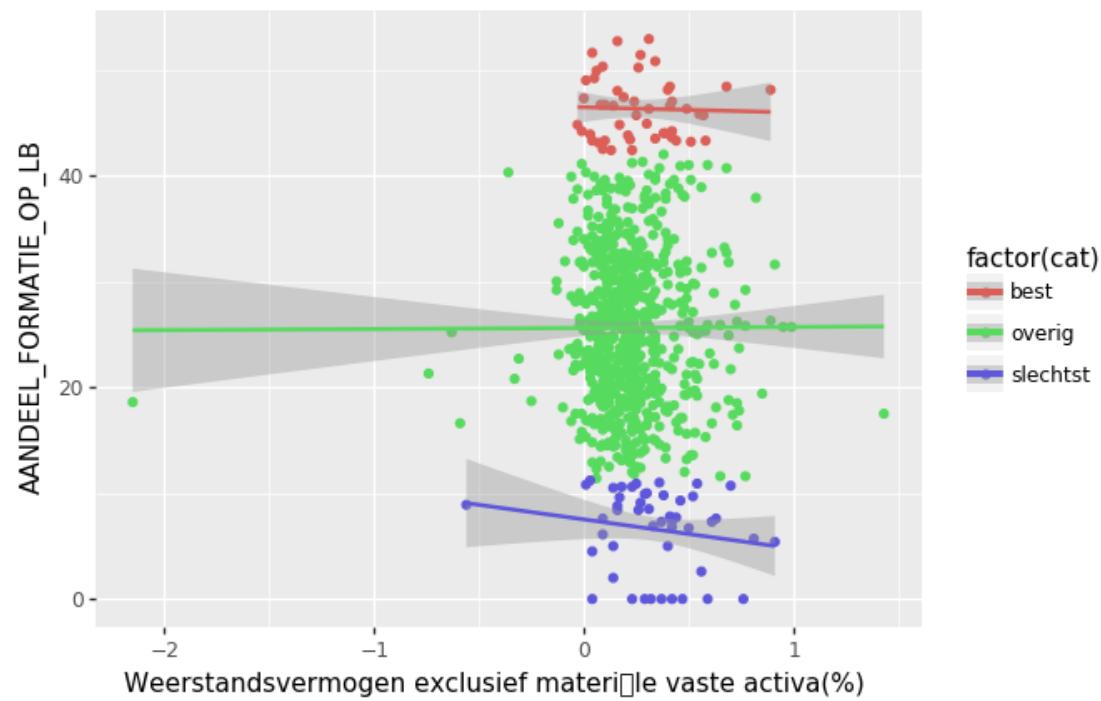


OMVANG\_FORMATIE\_LIO (804 rows, 0.19/-0.02/0.02 alles 0.01)

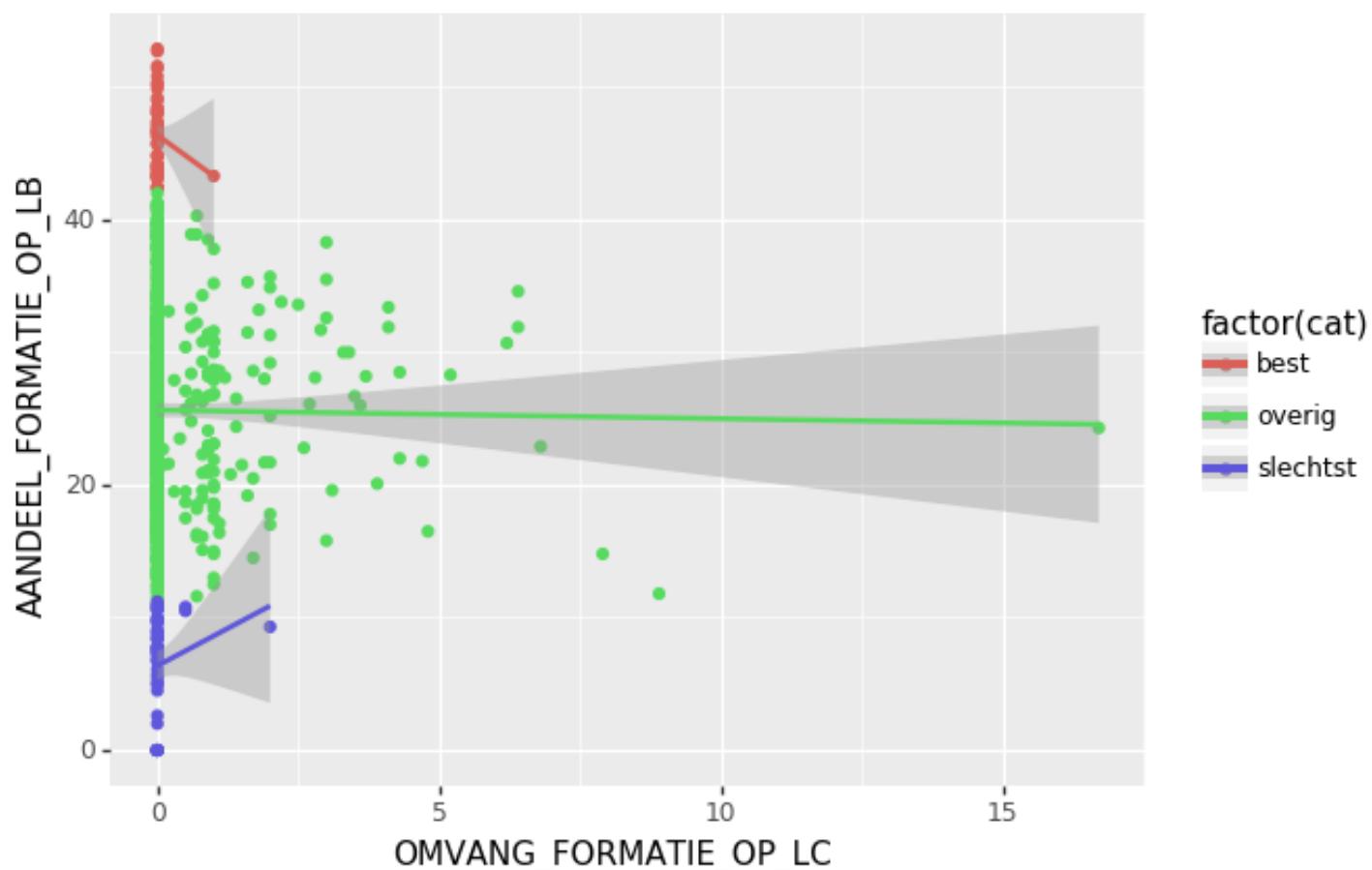




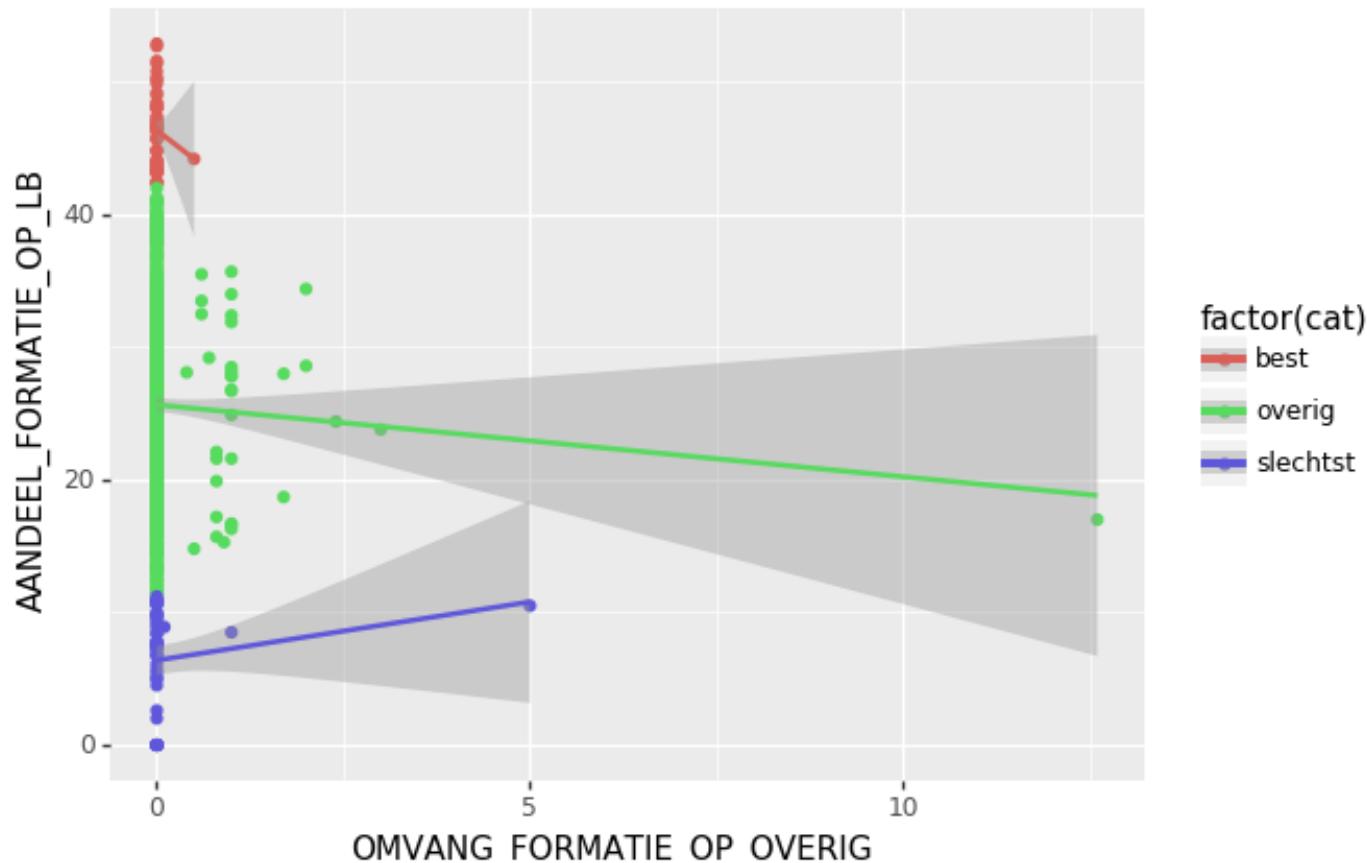
Weerstandsvermogen exclusief materiële vaste activa(%) (791 rows, -0.03/0.00/-0.18 alles -0.04)



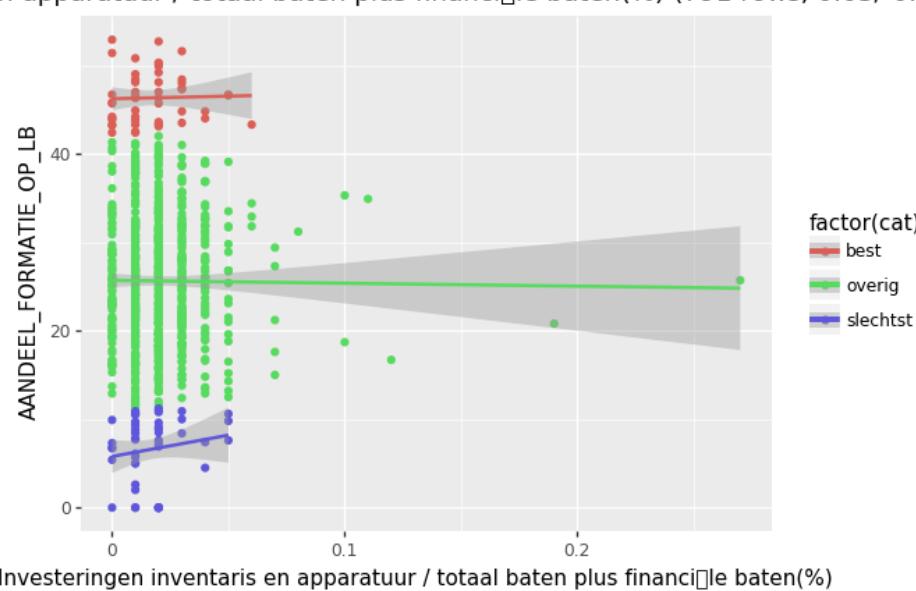
OMVANG\_FORMATIE\_OP\_LC (804 rows, -0.15/-0.01/0.17 alles -0.01)



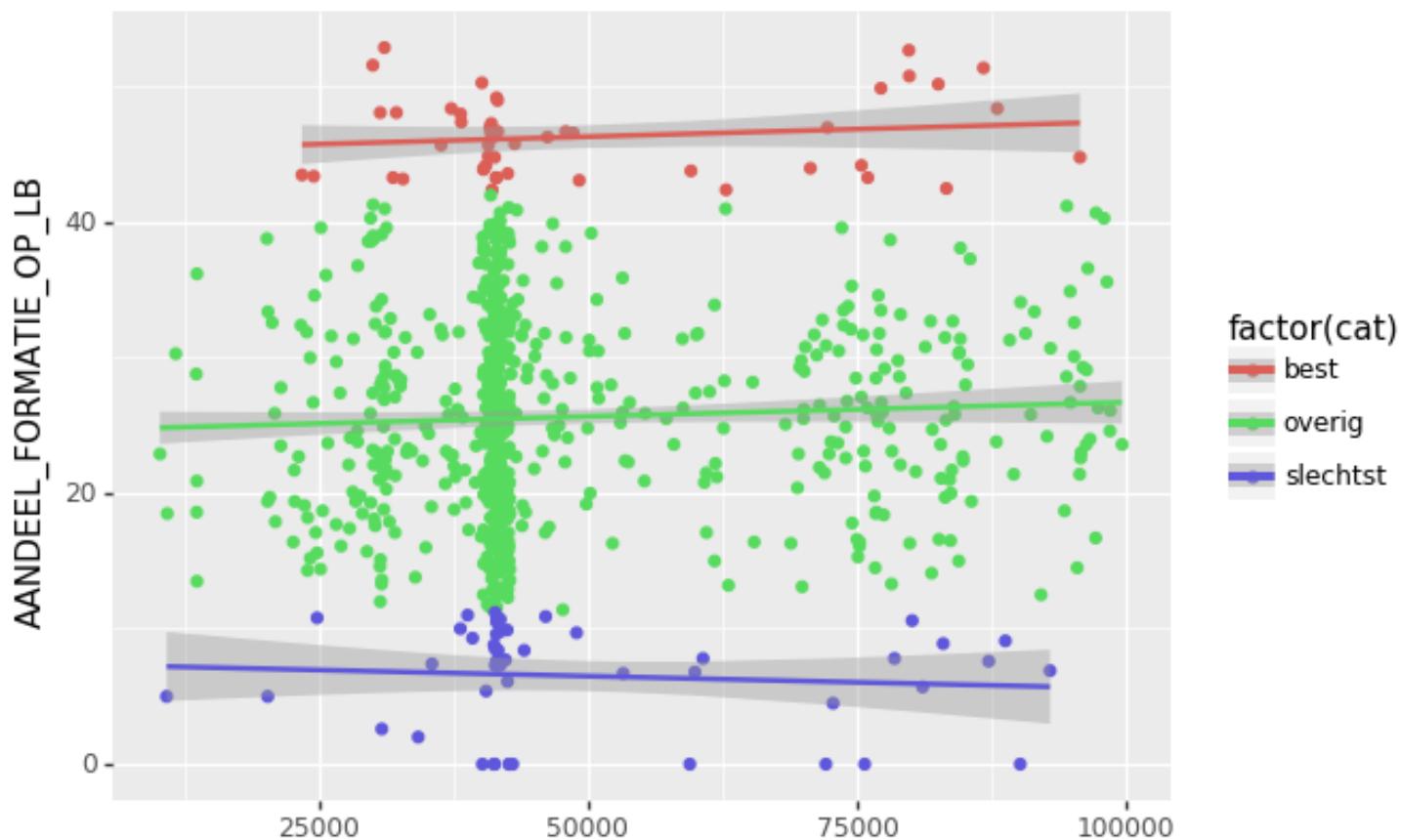
## OMVANG\_FORMATIE\_OP\_OVERIG (804 rows, -0.11/-0.04/0.16 alles -0.05)



## Investeringen inventaris en apparatuur / totaal baten plus financiële baten(%) (791 rows, 0.03/-0.01/0.16 alles -0.00)

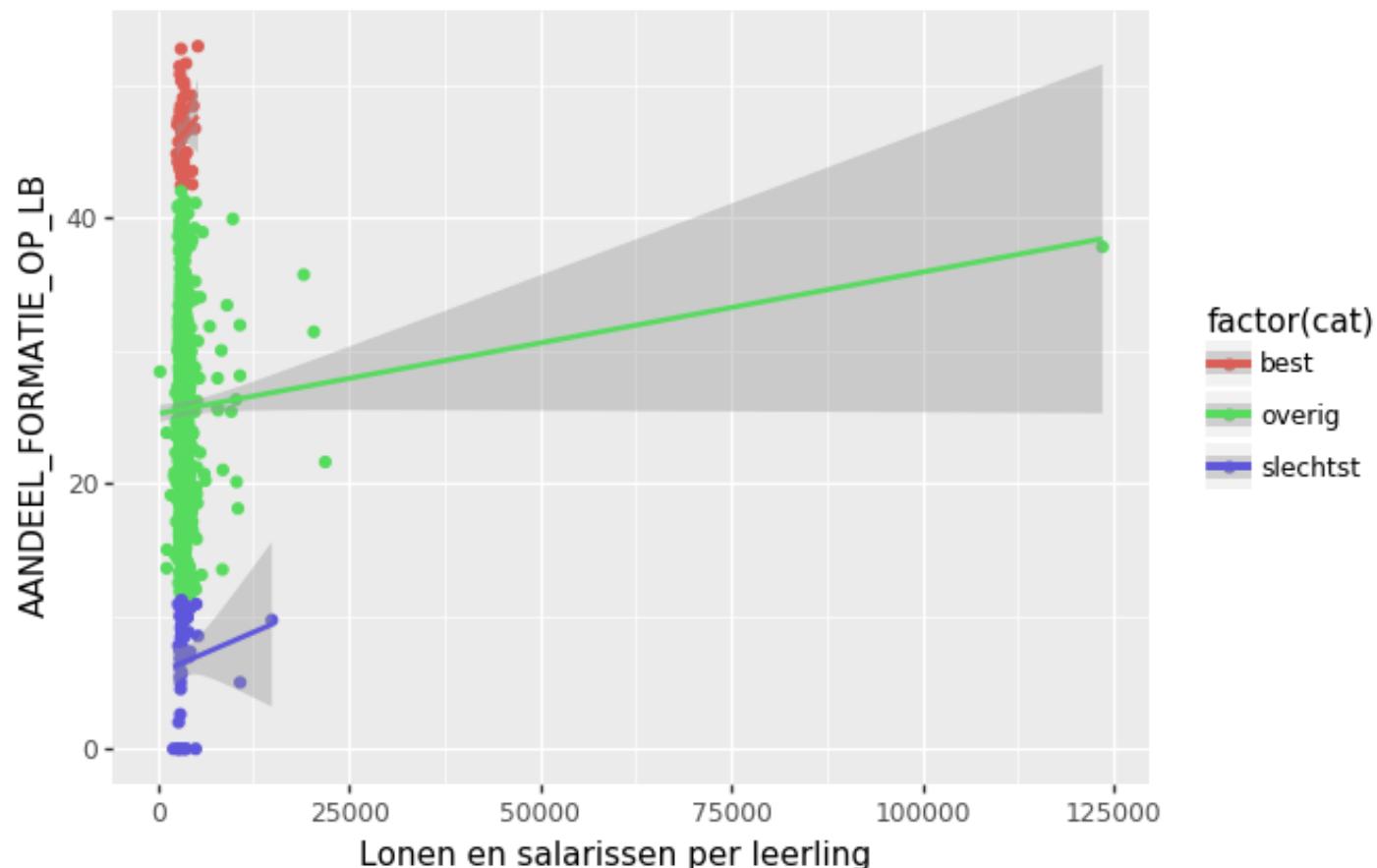


## BEVOEGD\_GEZAGNUMMER (804 rows, 0.15/0.05/-0.09 alles 0.04)

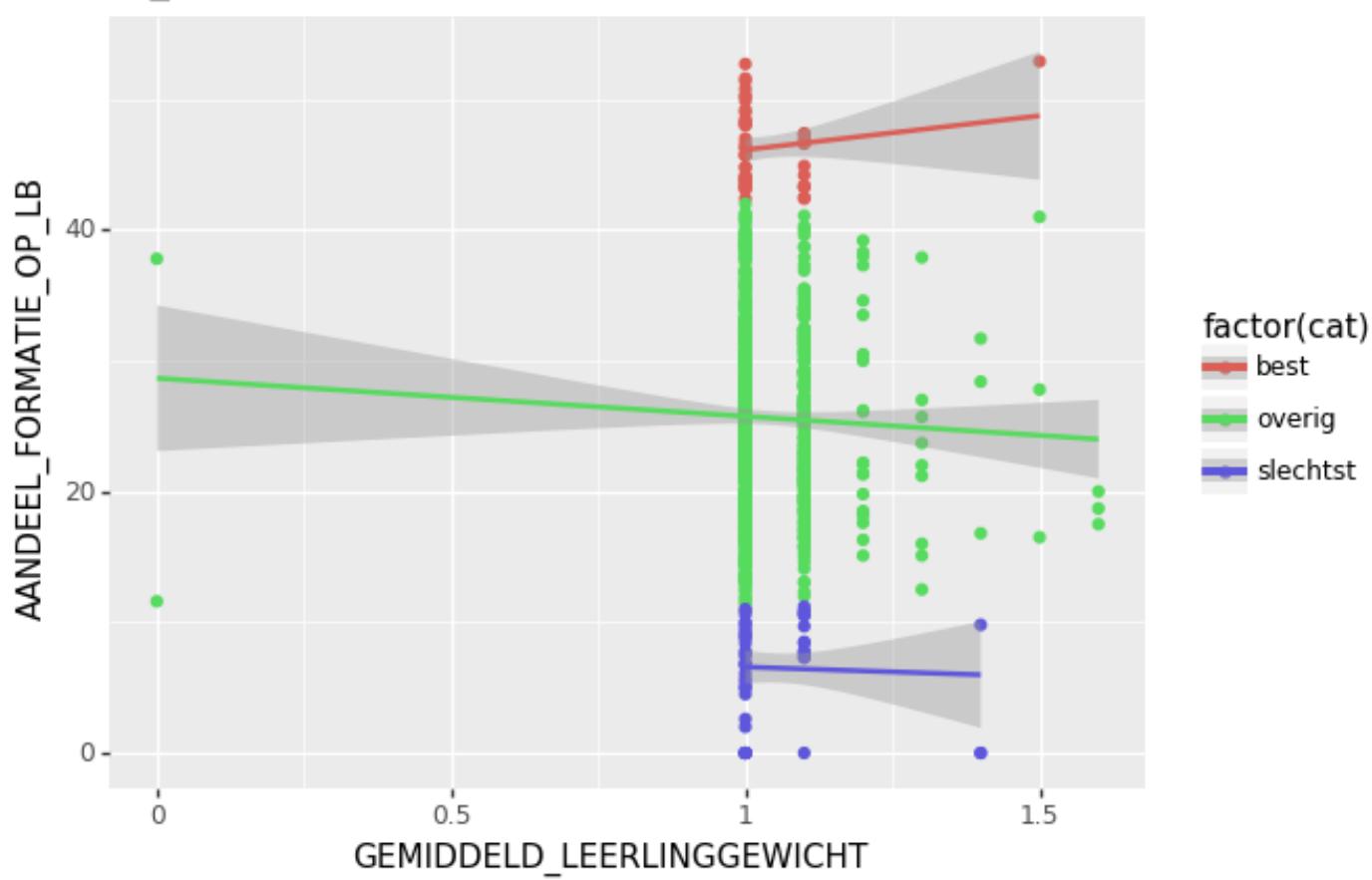


# BEVOEGD\_GEZAGNUMMER

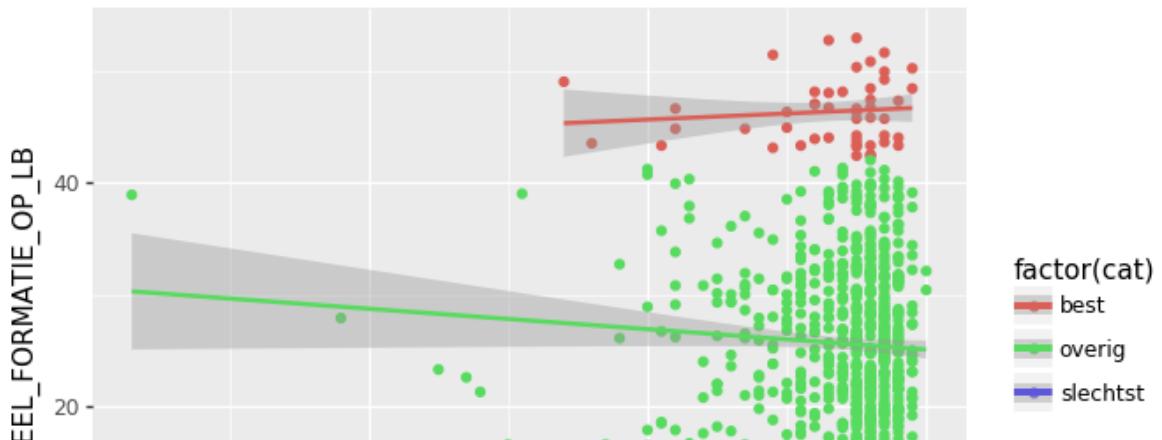
Lonen en salarissen per leerling (791 rows, 0.15/0.07/0.13 alles 0.04)

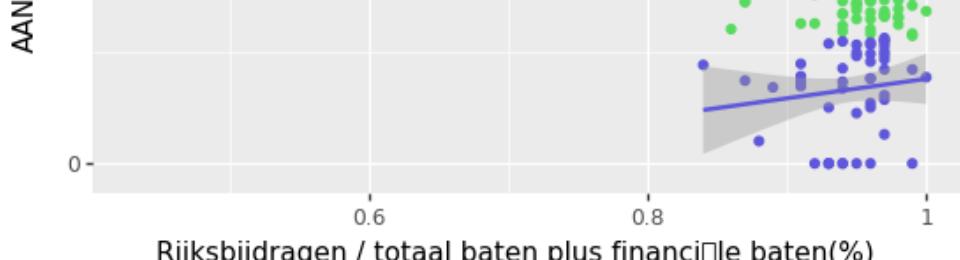


GEMIDDELD\_LEERLINGGEWICHT (804 rows, 0.15/-0.04/-0.04 alles -0.05)

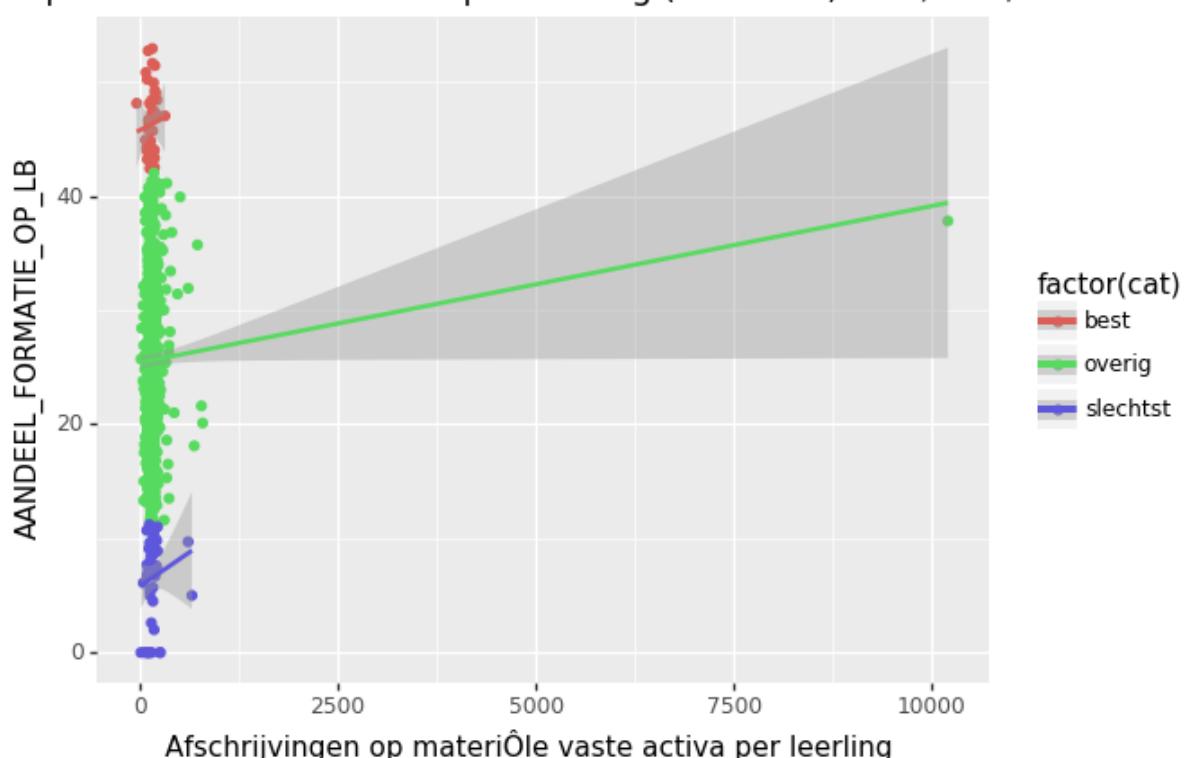


Rijksbijdragen / totaal baten plus financiële baten(%) (791 rows, 0.10/-0.07/0.14 alles -0.08)

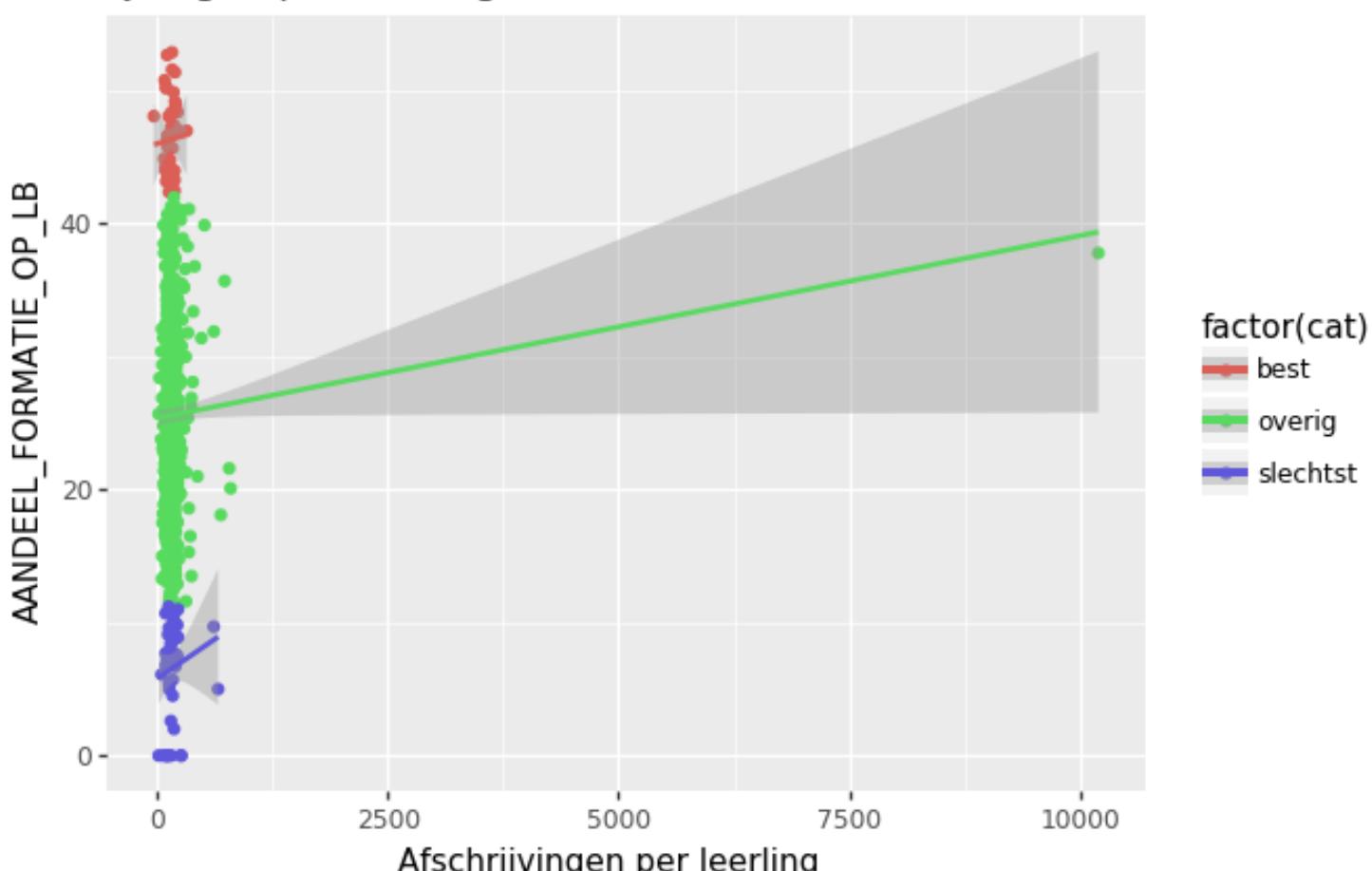




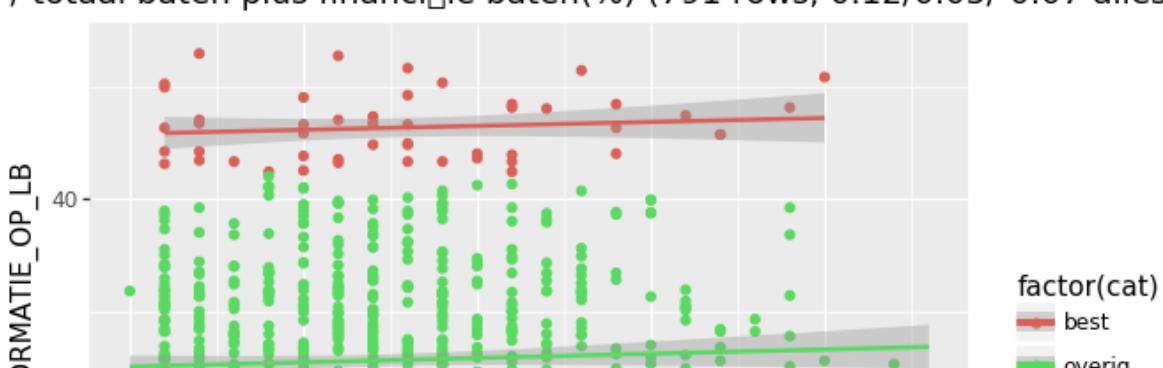
Afschrijvingen op materiële vaste activa per leerling (789 rows, 0.06/0.08/0.14 alles 0.04)

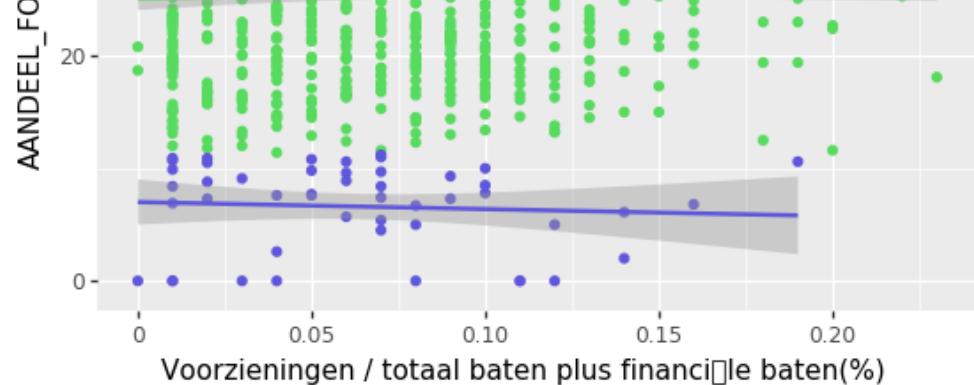


Afschrijvingen per leerling (791 rows, 0.04/0.08/0.14 alles 0.04)



Voorzieningen / totaal baten plus financiële baten(%) (791 rows, 0.12/0.05/-0.07 alles 0.07)

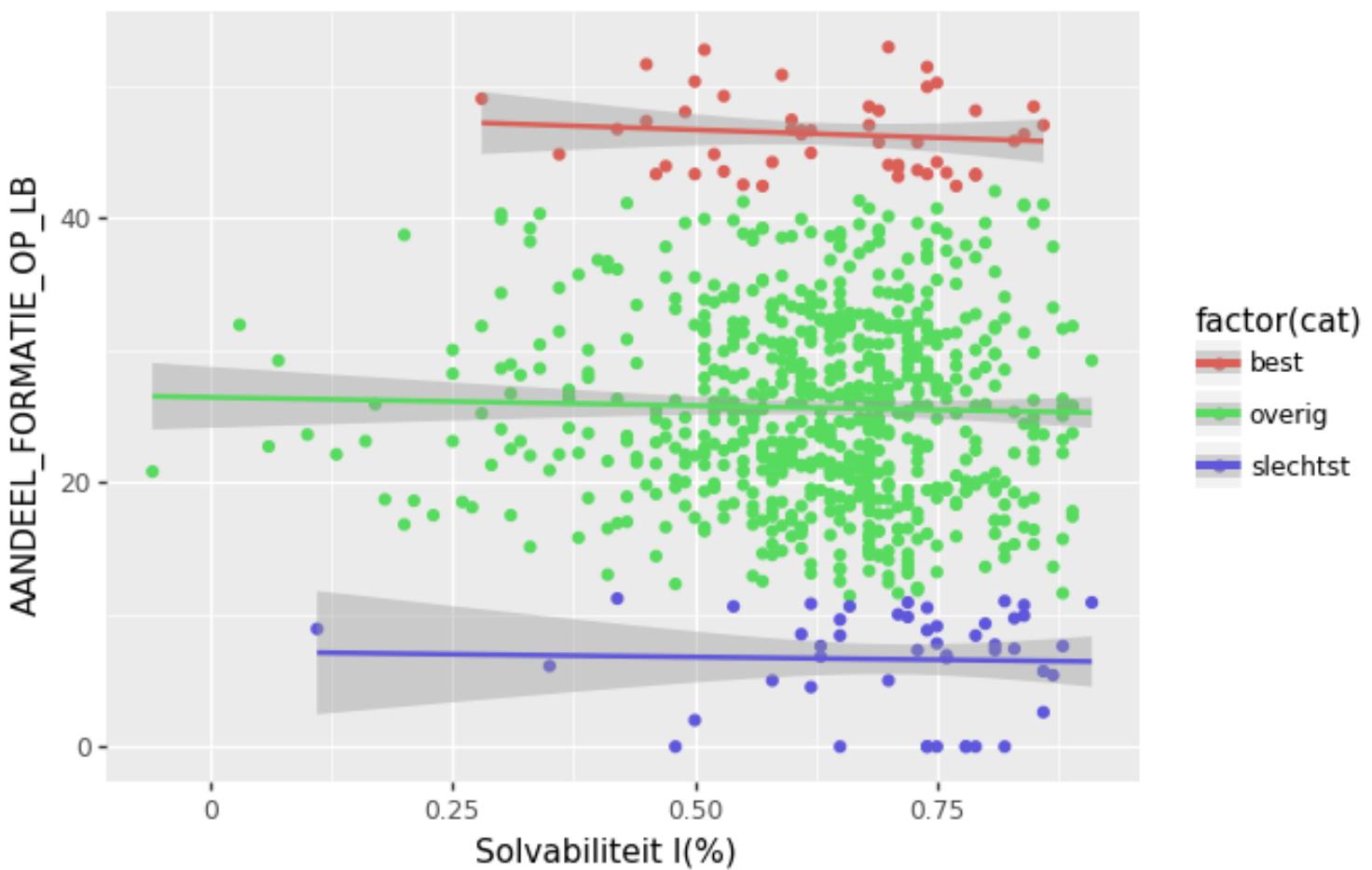




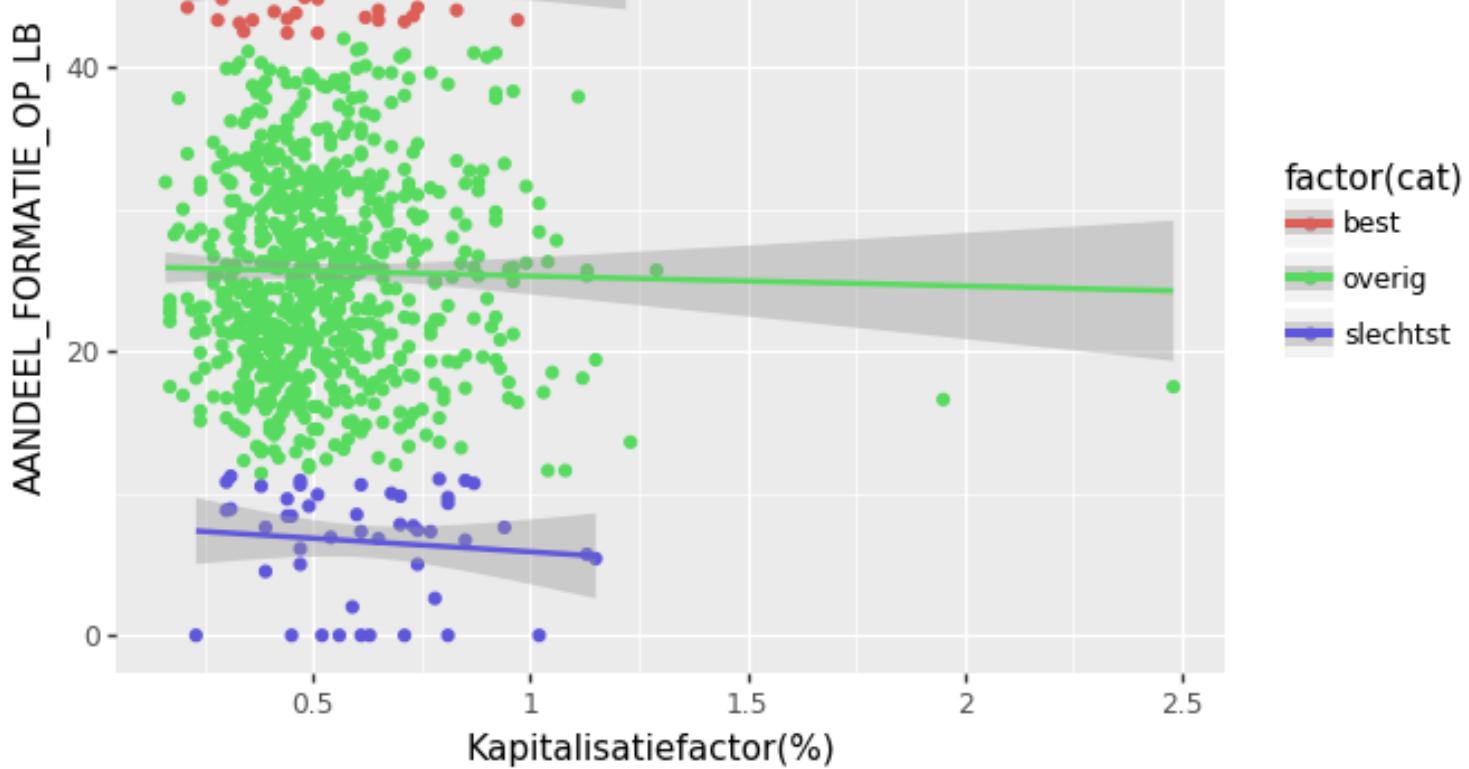
Schoonmaakkosten per leerling (787 rows, -0.11/0.11/0.09 alles 0.07)



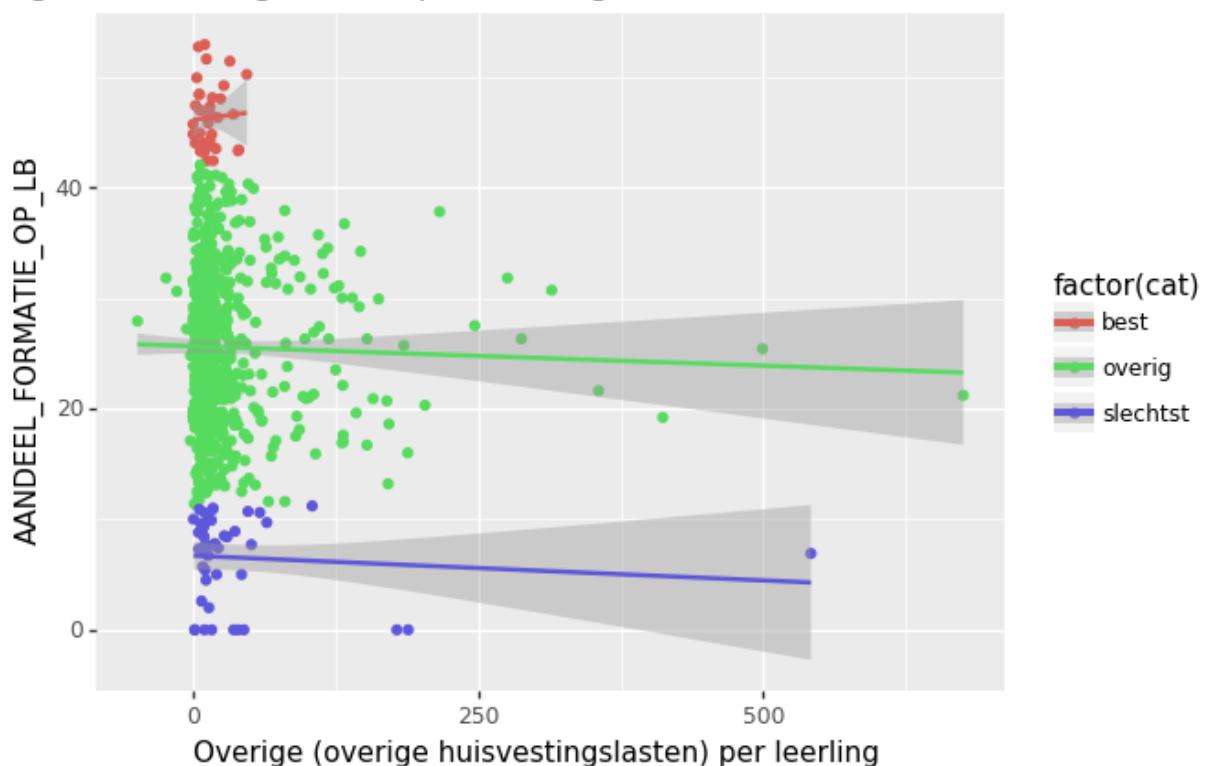
Solvabiliteit I(%) (791 rows, -0.11/-0.03/-0.03 alles -0.08)



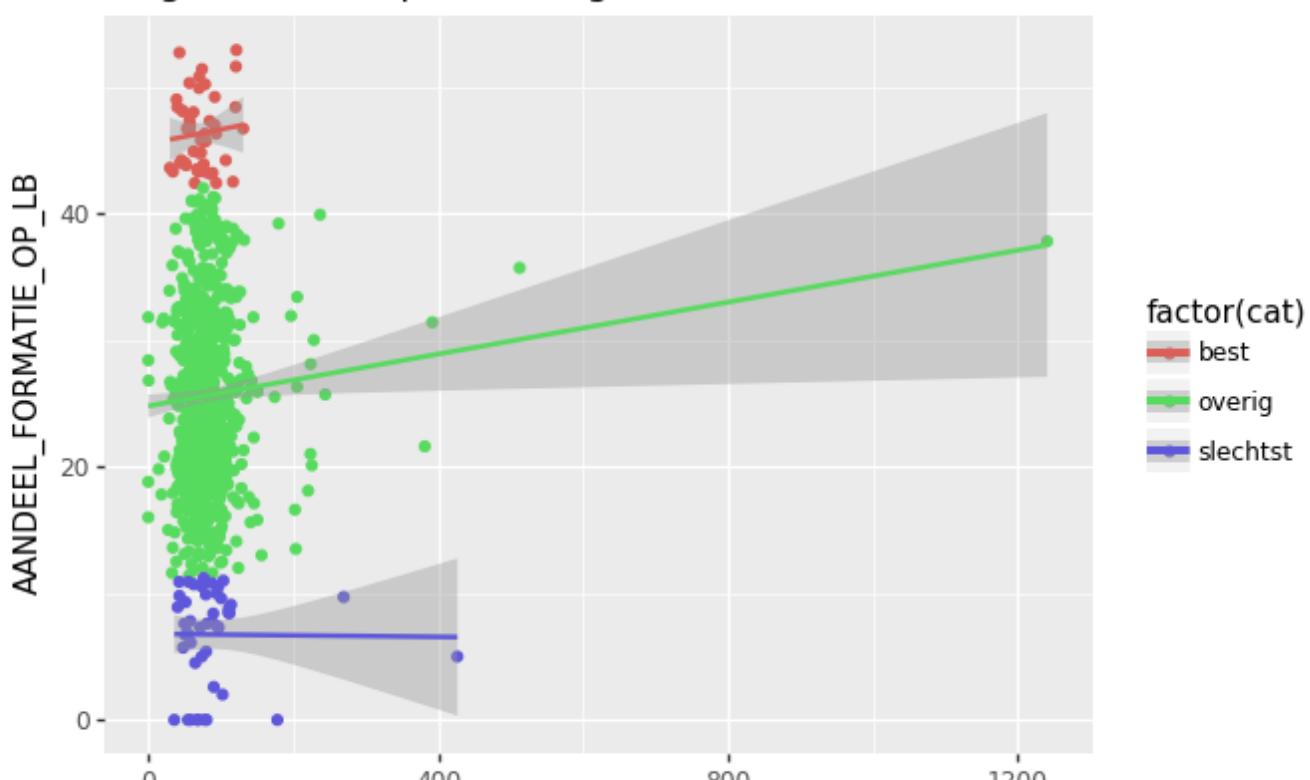
Kapitalisatiefactor(%) (791 rows, 0.03/-0.02/-0.11 alles -0.04)



Overige (overige huisvestingslasten) per leerling (703 rows, 0.05/-0.03/-0.10 alles -0.07)

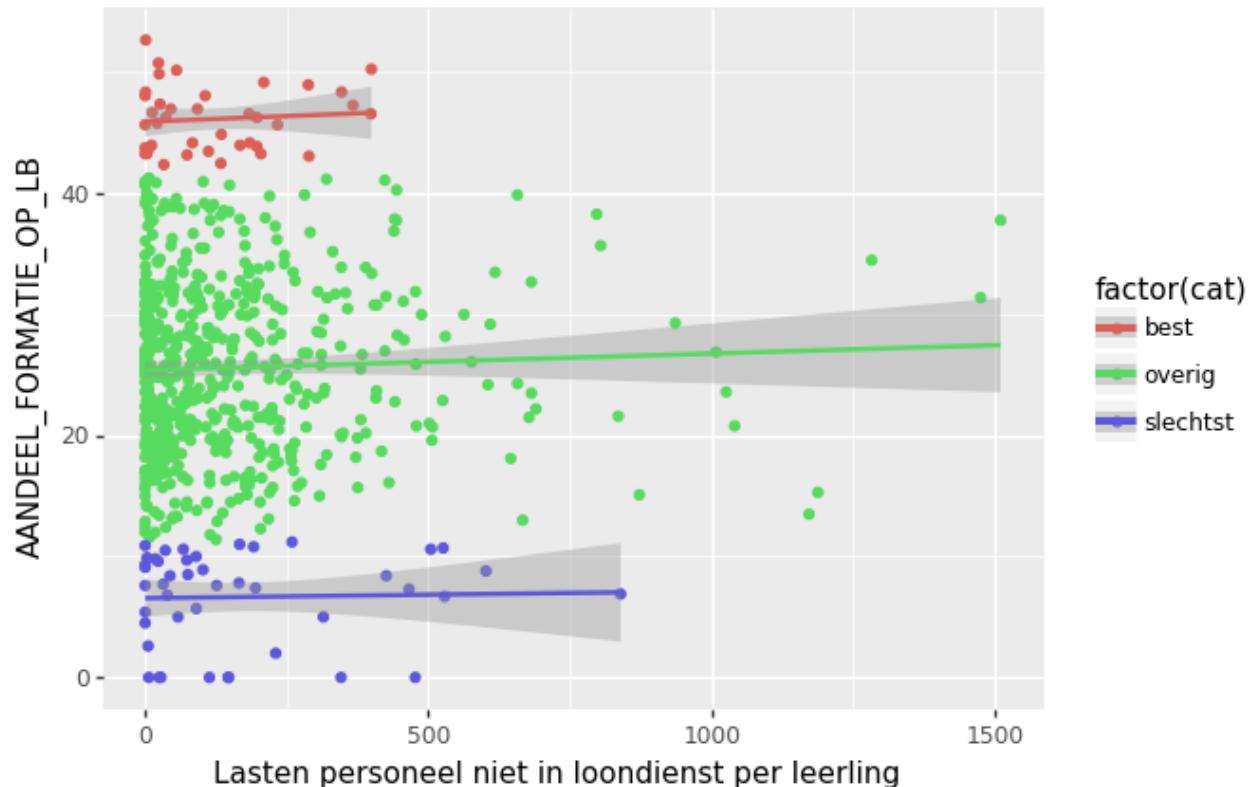


Lasten voor energie en water per leerling (785 rows, 0.09/0.09/-0.01 alles 0.02)

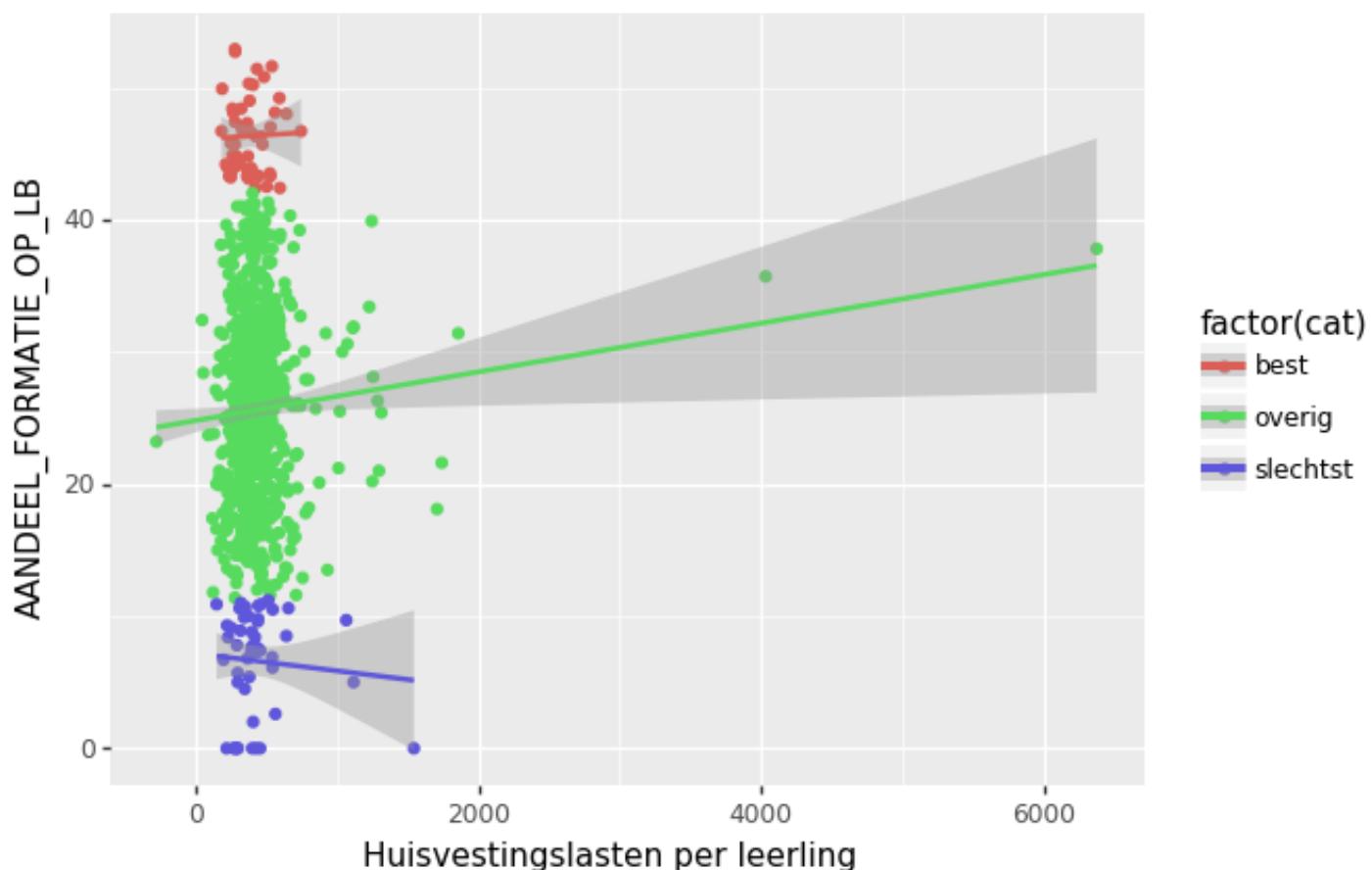


### Lasten voor energie en water per leerling

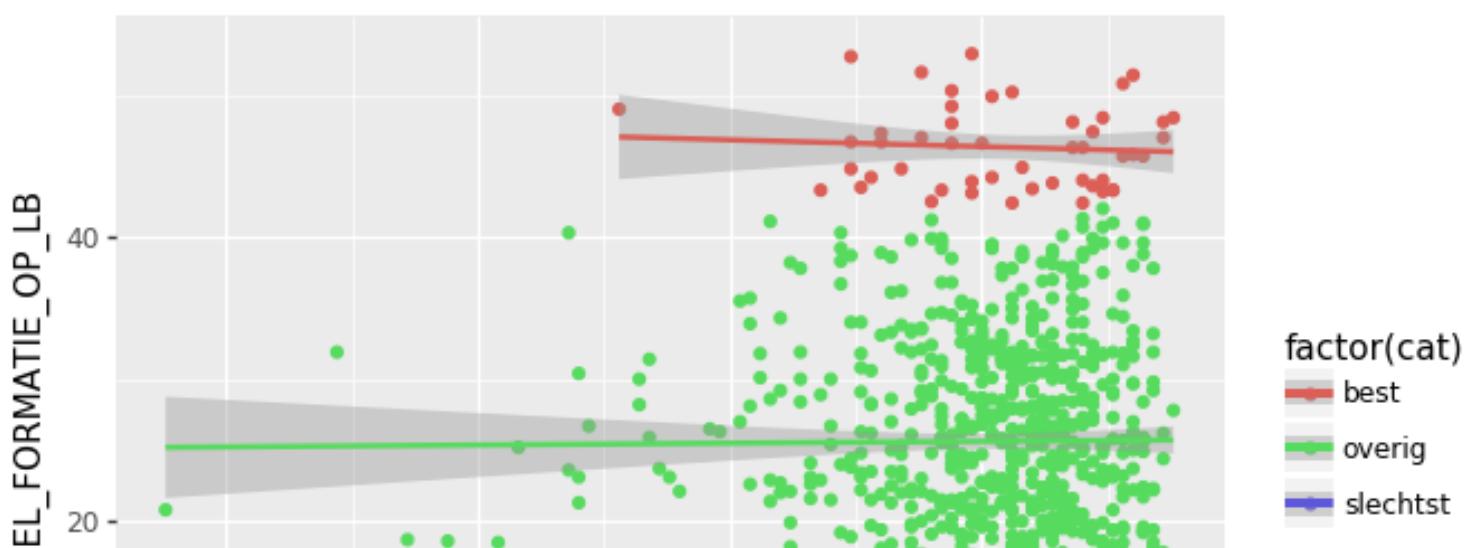
Lasten personeel niet in loondienst per leerling (665 rows, 0.09/0.04/0.03 alles -0.01)



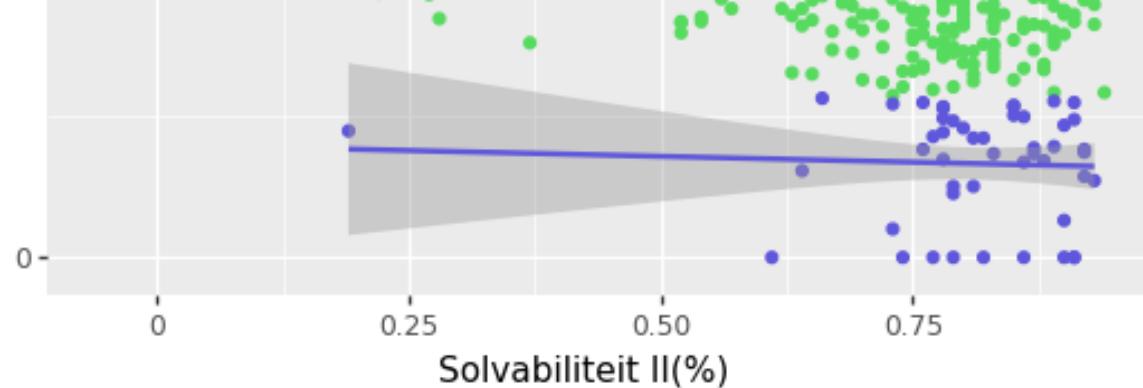
Huisvestingslasten per leerling (791 rows, 0.03/0.08/-0.08 alles 0.03)



Solvabiliteit II(%) (791 rows, -0.07/0.01/-0.05 alles -0.03)



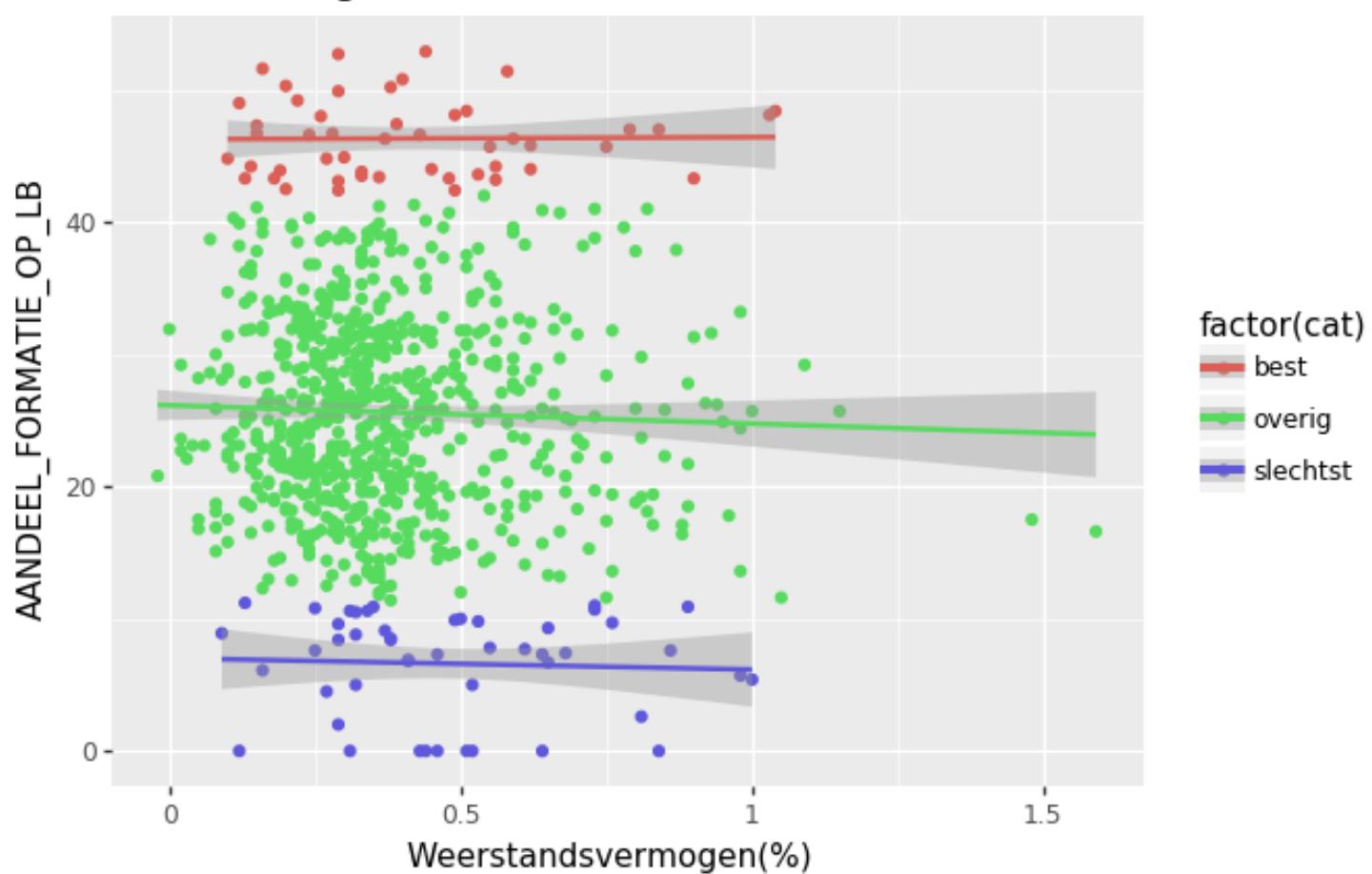
AANDE



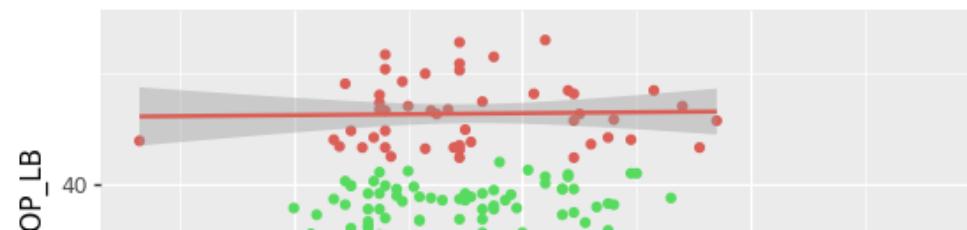
Onderhoudslasten (klein onderhoud) per leerling (782 rows, -0.03/0.02/-0.07 alles -0.00)

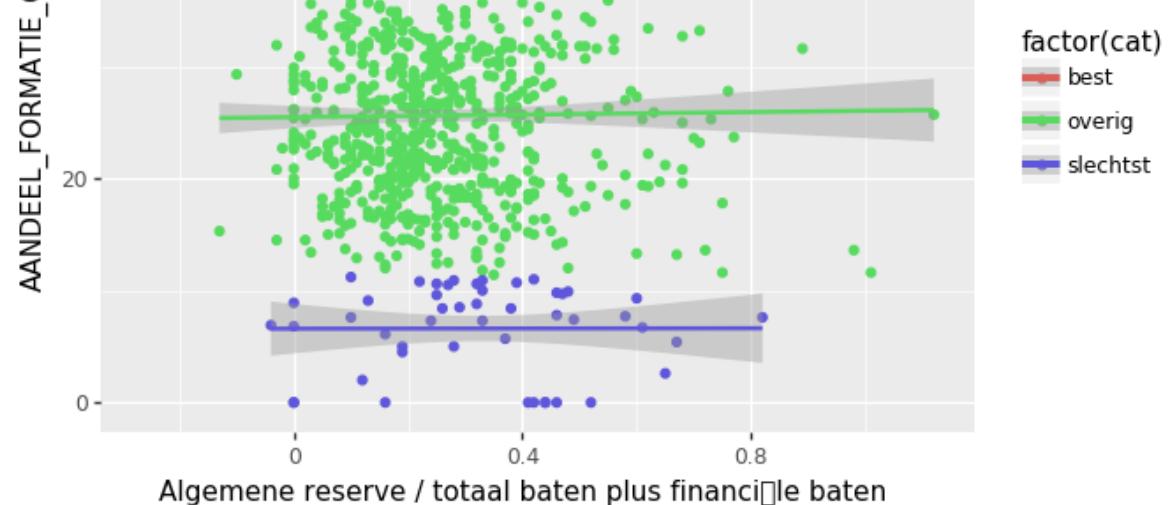


Weerstandsvermogen(%) (791 rows, 0.01/-0.04/-0.05 alles -0.06)

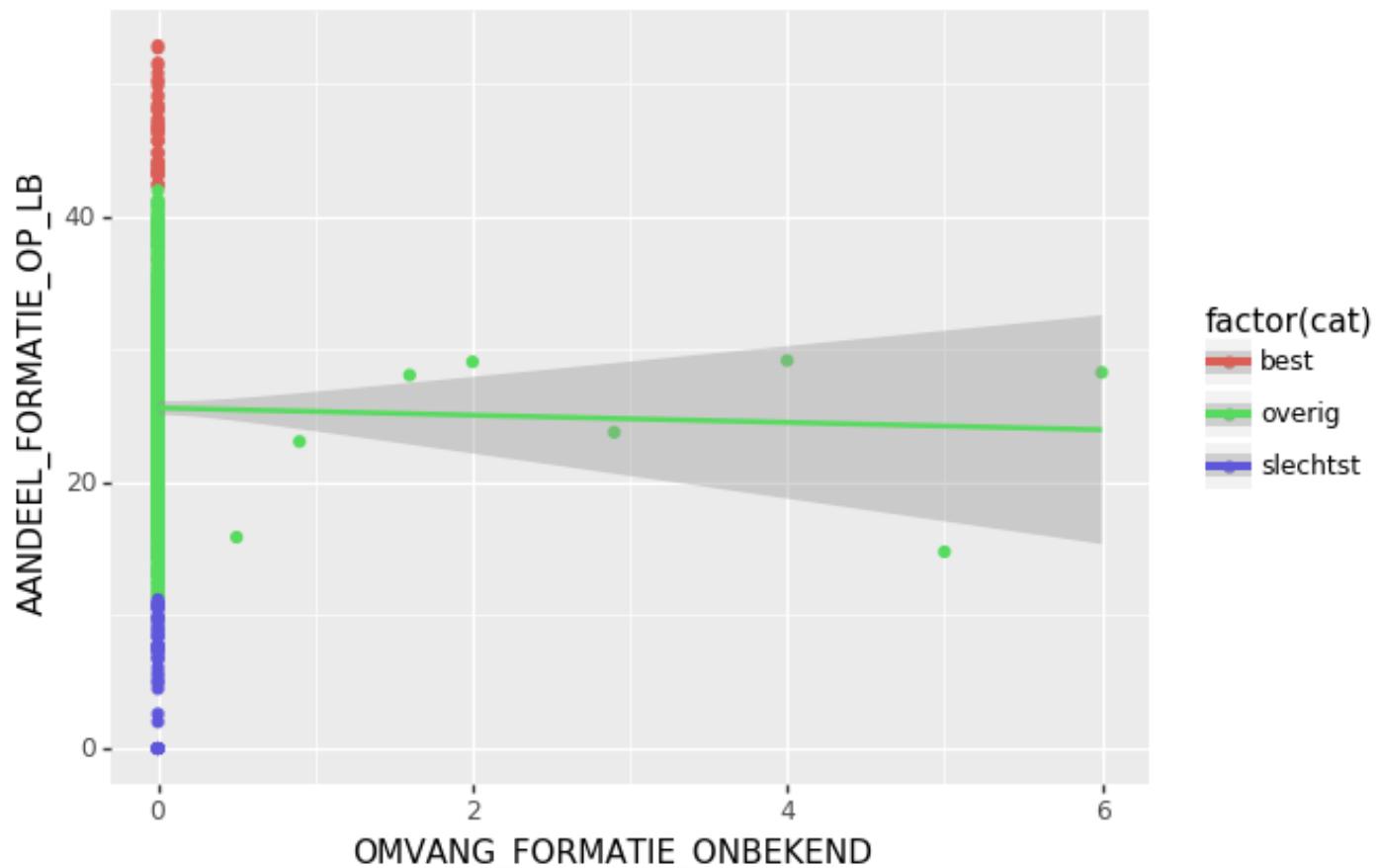


Algemene reserve / totaal baten plus financiële baten (791 rows, 0.03/0.01/0.00 alles -0.00)

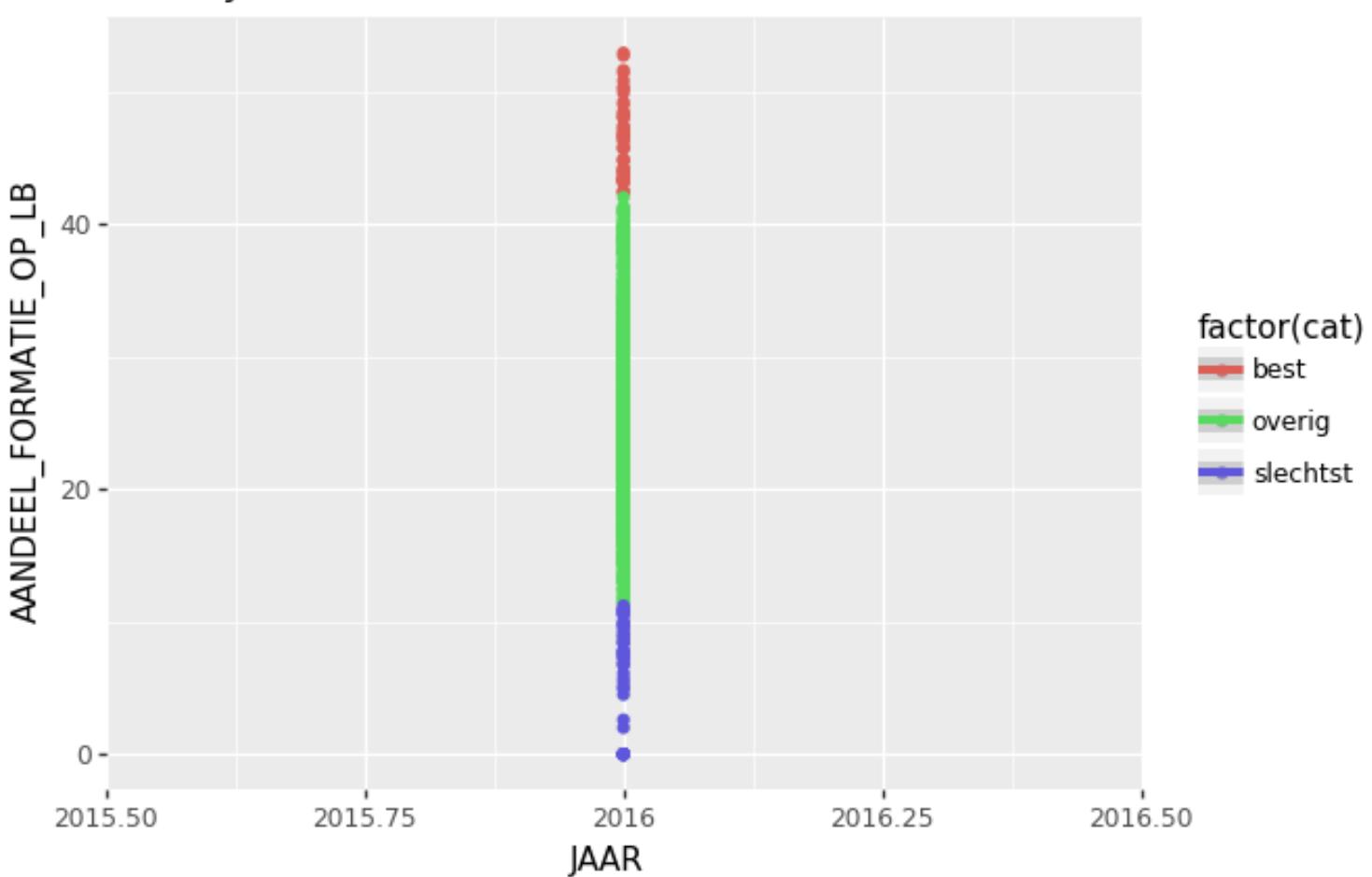




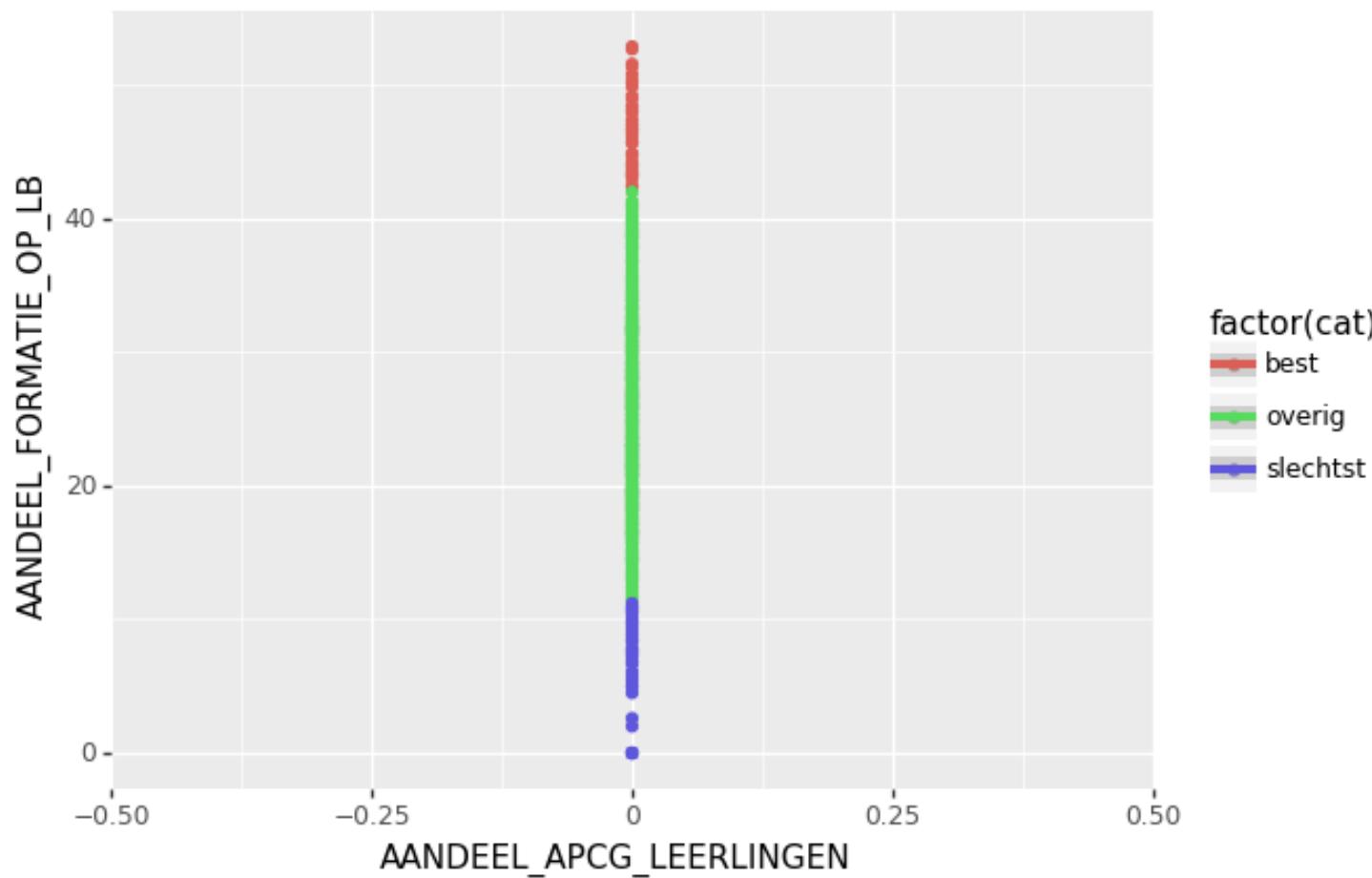
OMVANG\_FORMATIE\_ONBEKEND (804 rows, nan/-0.01/nan alles -0.01)



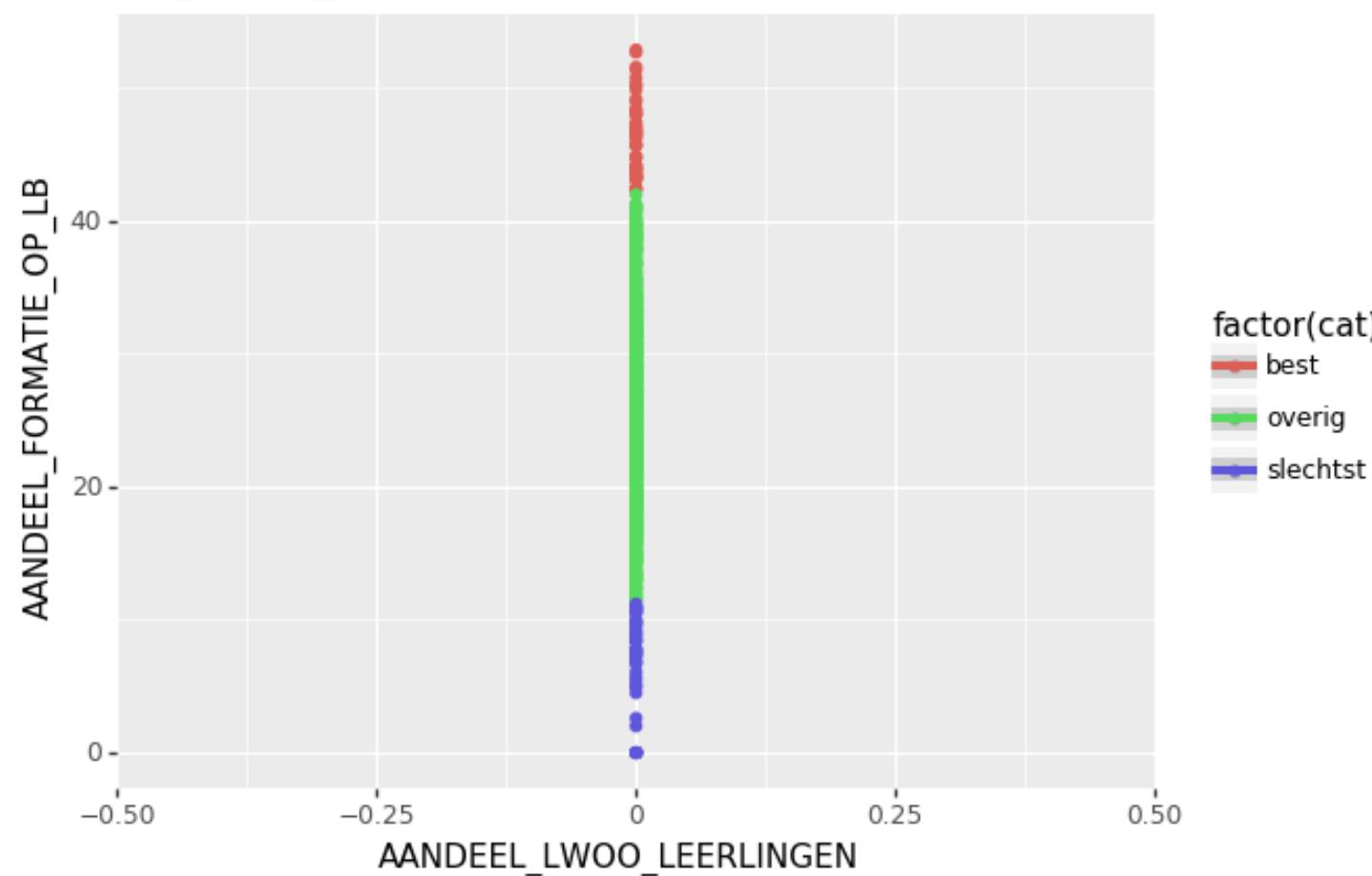
JAAR (804 rows, nan/nan/nan alles nan)



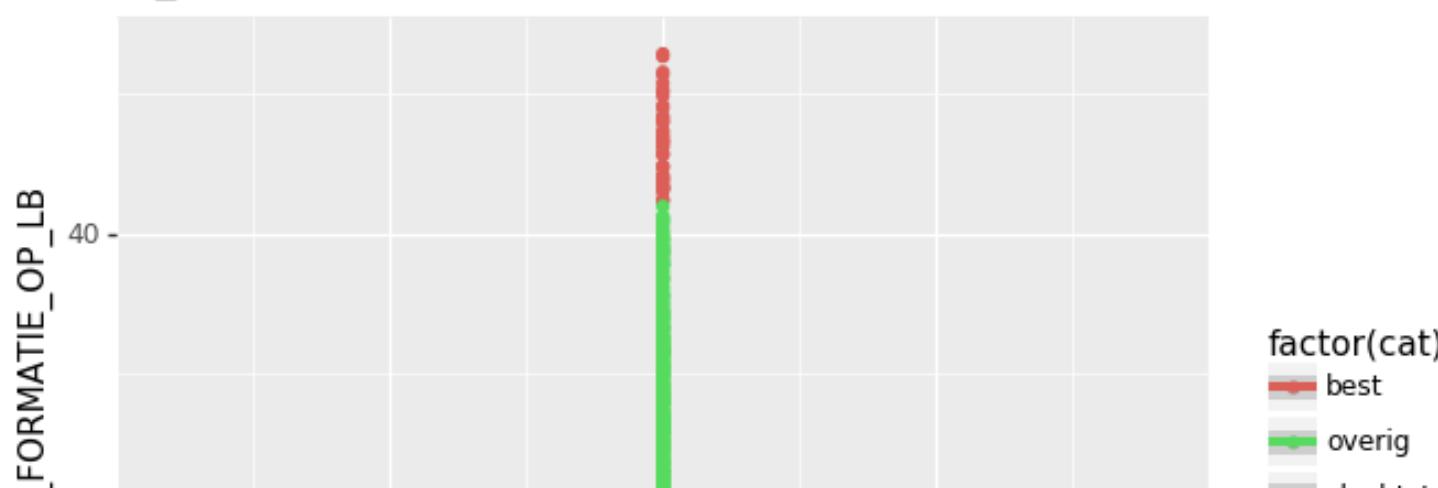
AANDEEL\_APCG\_LEERLINGEN (804 rows, nan/nan/nan alles nan)

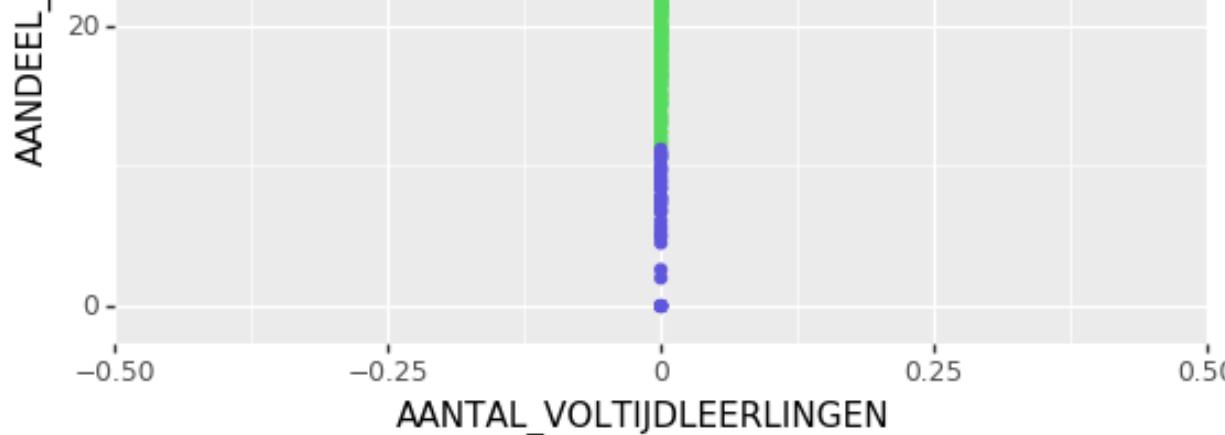


AANDEEL\_LWOO\_LEERLINGEN (804 rows, nan/nan/nan alles nan)

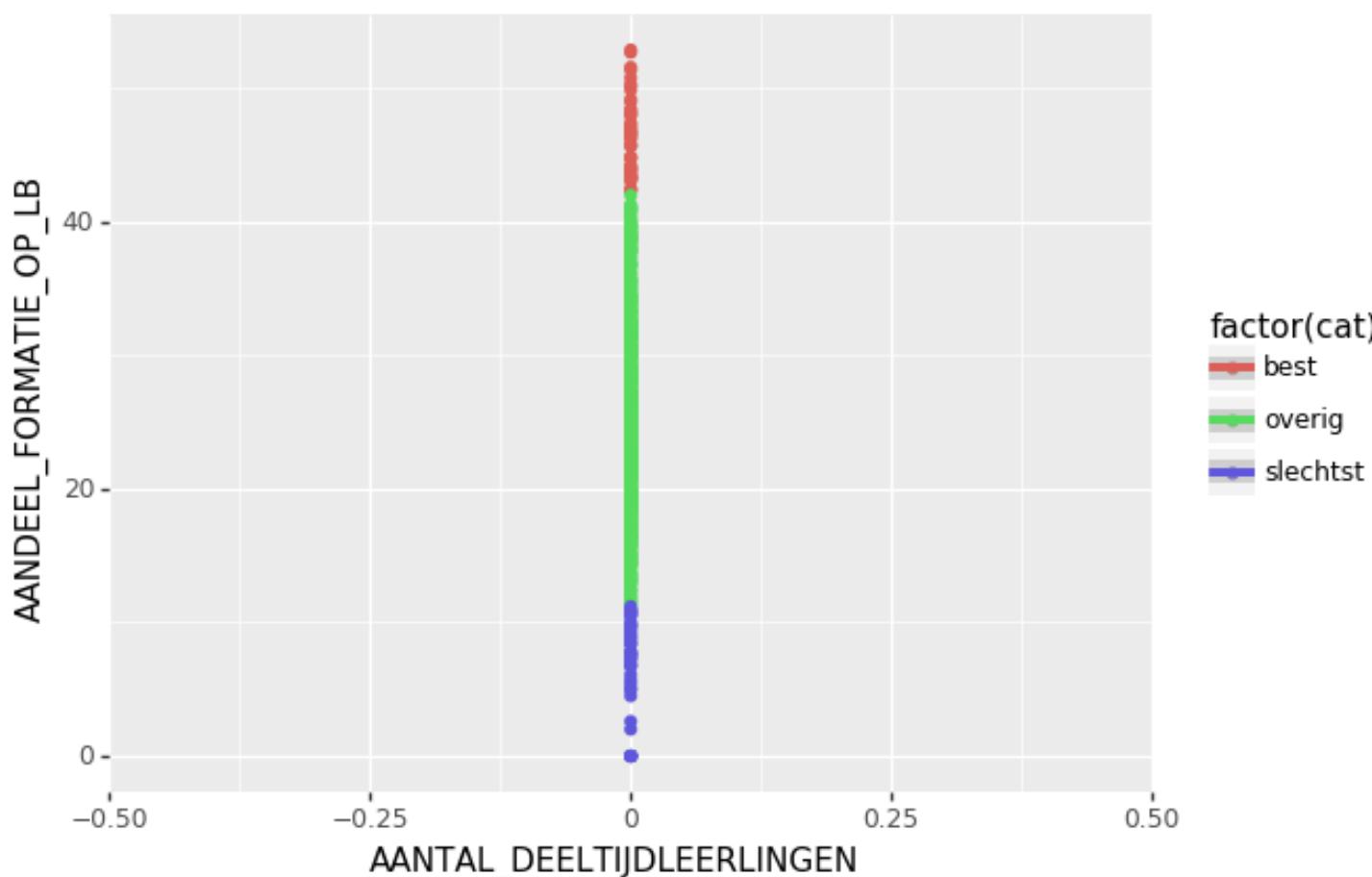


AANTAL\_VOLTIJDLEERLINGEN (804 rows, nan/nan/nan alles nan)

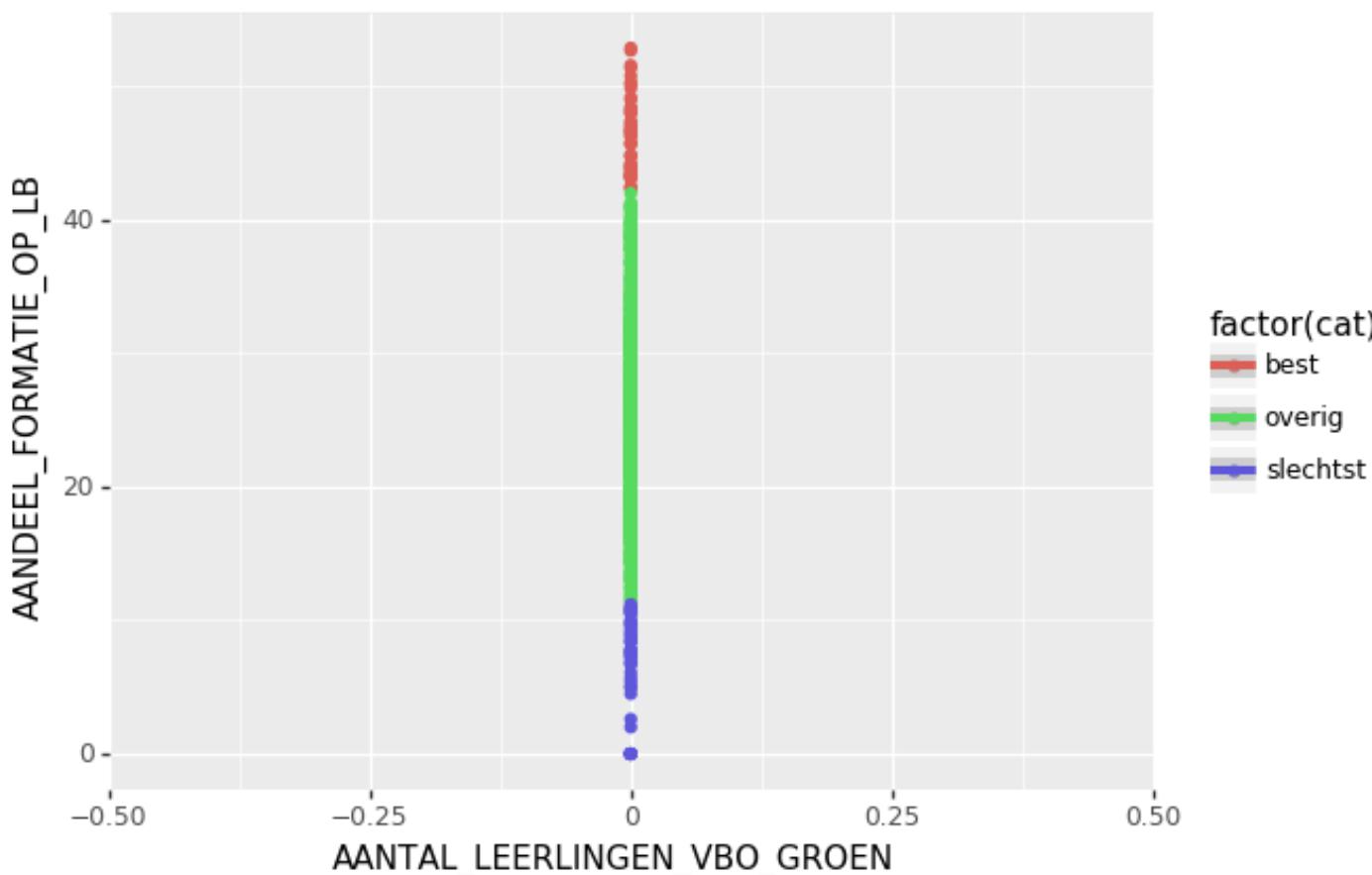




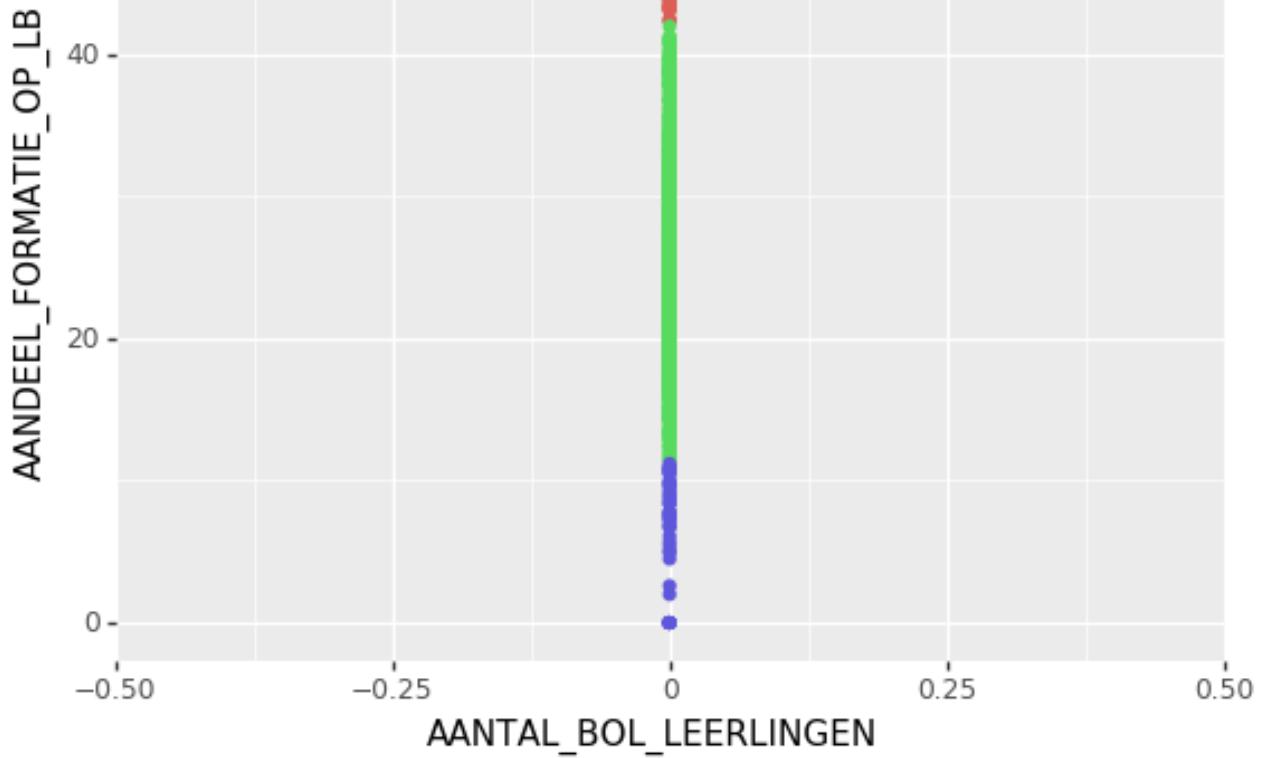
AANTAL\_DEELTIJDLEERLINGEN (804 rows, nan/nan/nan alles nan)



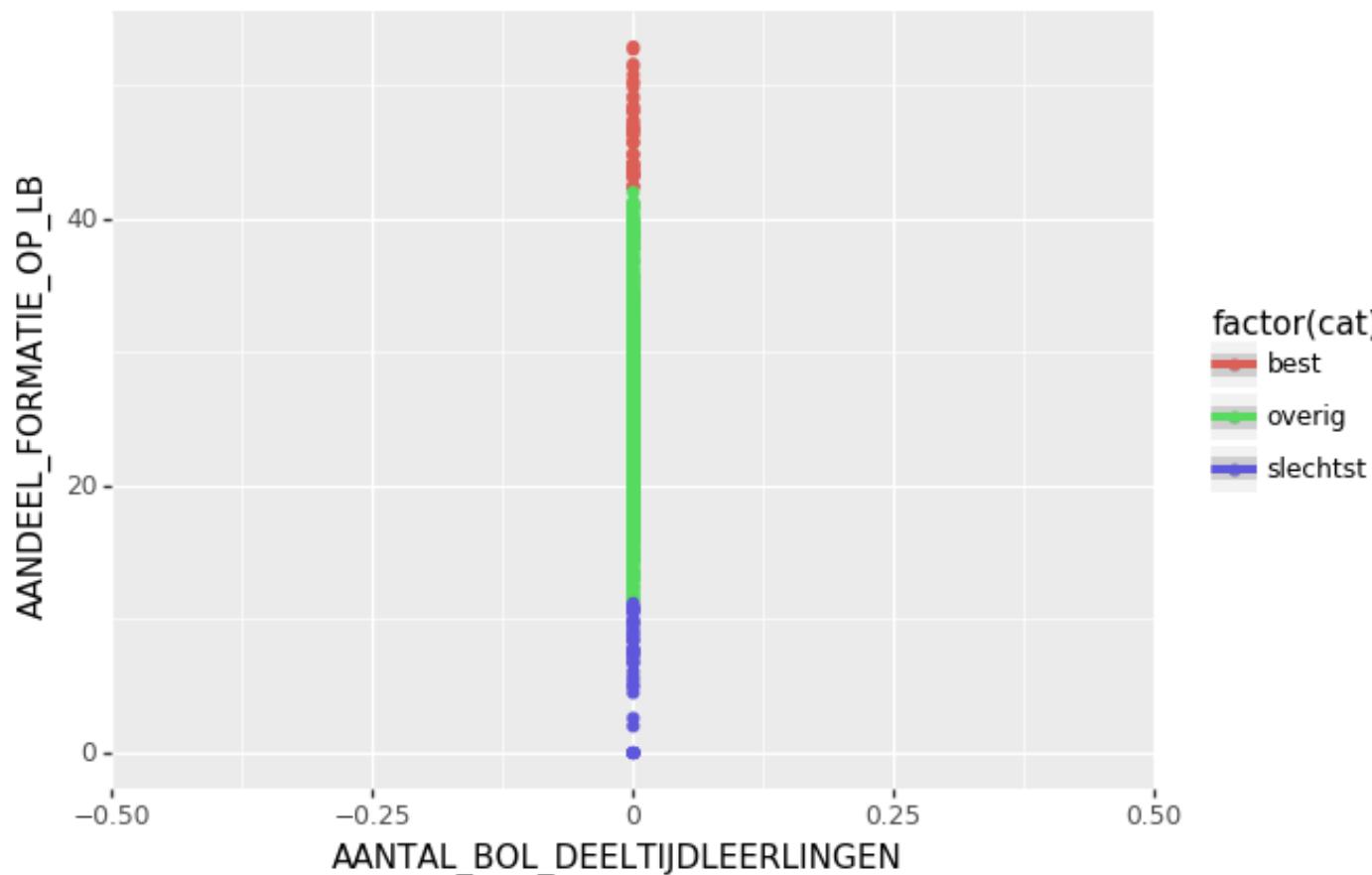
AANTAL\_LEERLINGEN\_VBO\_GROEN (804 rows, nan/nan/nan alles nan)



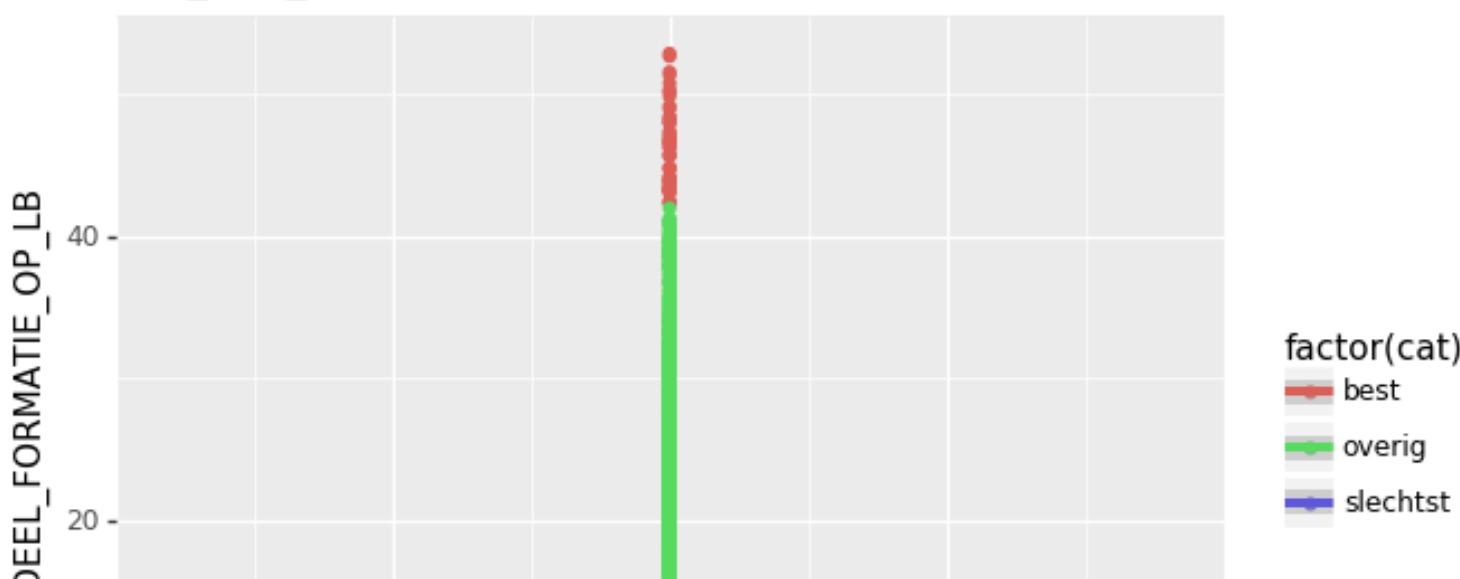
AANTAL\_BOL\_LEERLINGEN (804 rows, nan/nan/nan alles nan)

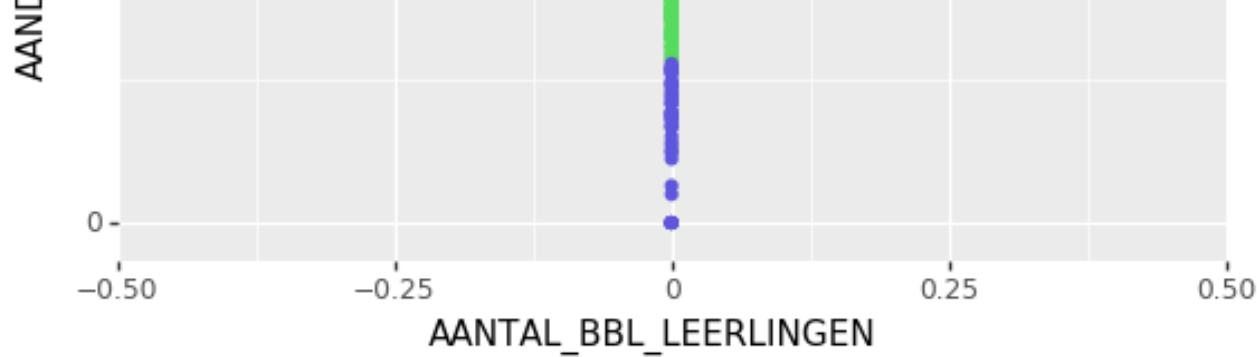


AANTAL\_BOL\_DEELTIJDLEERLINGEN (804 rows, nan/nan/nan alles nan)

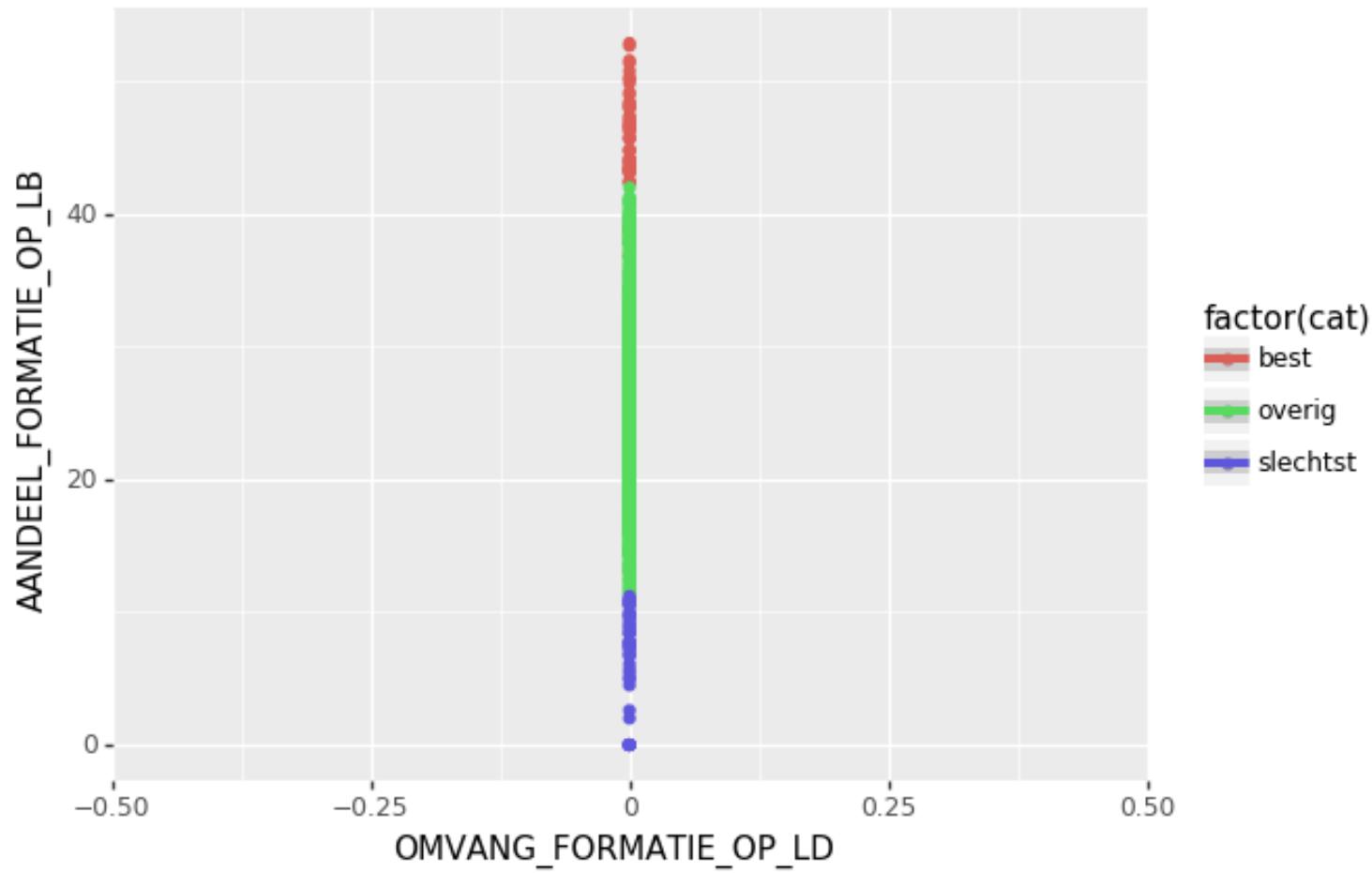


AANTAL\_BBL\_LEERLINGEN (804 rows, nan/nan/nan alles nan)

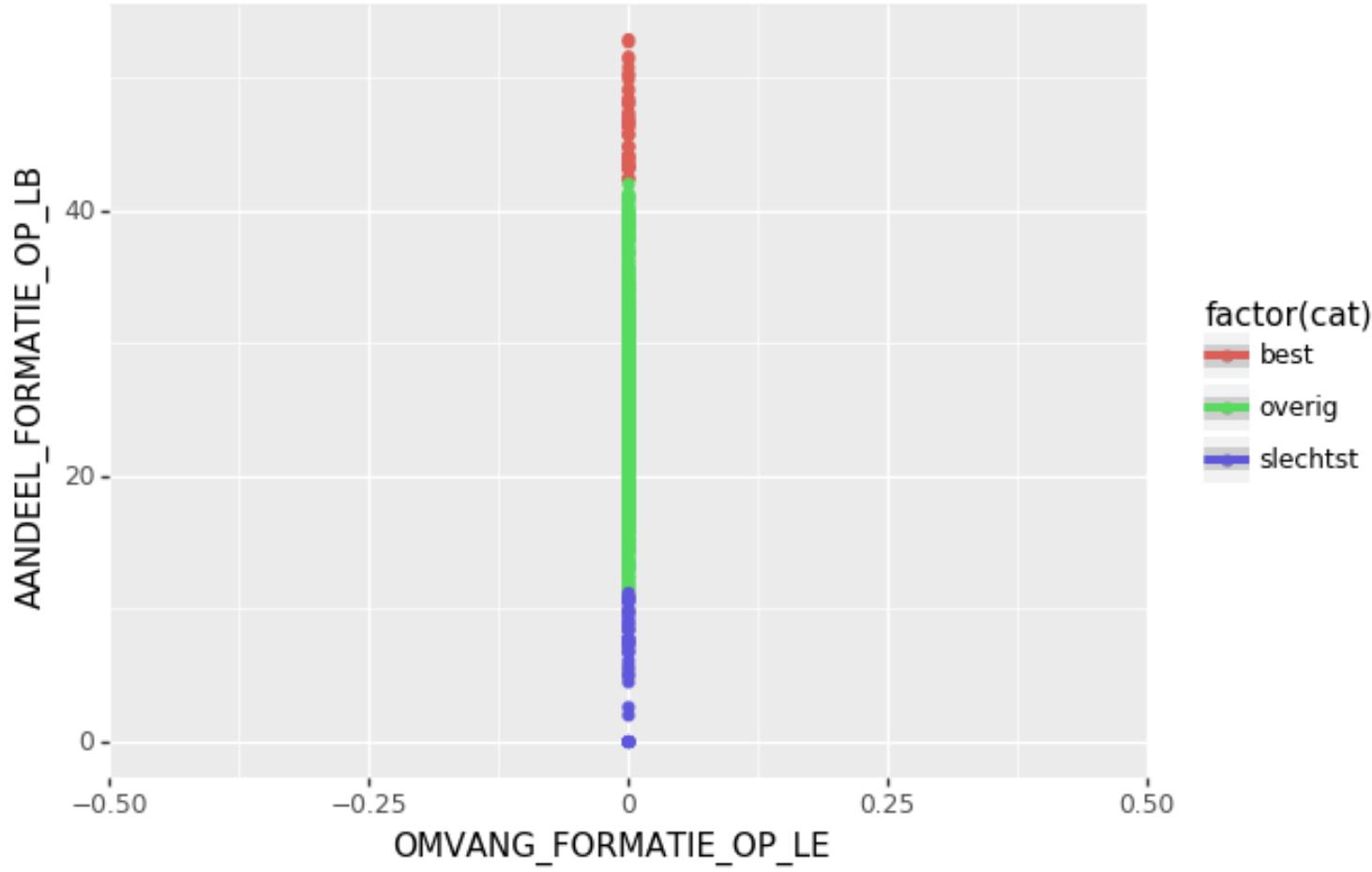




OMVANG\_FORMATIE\_OP\_LD (804 rows, nan/nan/nan alles nan)



OMVANG\_FORMATIE\_OP\_LE (804 rows, nan/nan/nan alles nan)



Out[14]:

[<ggplot>: 287914760]

```
[<ggplot: (287514700)>,
<ggplot: (287170680)>,
<ggplot: (-9223372036566151876)>,
<ggplot: (287899561)>,
<ggplot: (278714646)>,
<ggplot: (287911544)>,
<ggplot: (278702961)>,
<ggplot: (288613808)>,
<ggplot: (-9223372036566874017)>,
<ggplot: (288145258)>,
<ggplot: (-9223372036575153397)>,
<ggplot: (278716108)>,
<ggplot: (-9223372036566248879)>,
<ggplot: (288535002)>,
<ggplot: (-9223372036567149831)>,
<ggplot: (-9223372036575090926)>,
<ggplot: (-9223372036576060977)>,
<ggplot: (288618767)>,
<ggplot: (288519905)>,
<ggplot: (288517552)>,
<ggplot: (279672854)>,
<ggplot: (-9223372036566256089)>,
<ggplot: (-9223372036566860968)>,
<ggplot: (288534662)>,
<ggplot: (288611455)>,
<ggplot: (-9223372036566255272)>,
<ggplot: (-9223372036567617911)>,
<ggplot: (288622838)>,
<ggplot: (-9223372036575091389)>,
<ggplot: (288520470)>,
<ggplot: (-9223372036575091694)>,
<ggplot: (-9223372036575091249)>,
<ggplot: (279674474)>,
<ggplot: (288614839)>,
<ggplot: (288624048)>,
<ggplot: (288520796)>,
<ggplot: (287912519)>,
<ggplot: (288622698)>,
<ggplot: (279671963)>,
<ggplot: (-9223372036566864289)>,
<ggplot: (-9223372036575091519)>,
<ggplot: (278702863)>,
<ggplot: (278710855)>,
<ggplot: (288613647)>,
<ggplot: (286899243)>,
<ggplot: (-9223372036566161141)>,
<ggplot: (288937480)>,
<ggplot: (279685040)>,
<ggplot: (288534841)>,
<ggplot: (288522504)>,
<ggplot: (279673015)>,
<ggplot: (-9223372036566247350)>,
<ggplot: (-9223372036576061282)>,
<ggplot: (288622775)>,
<ggplot: (288527052)>,
<ggplot: (288528518)>,
<ggplot: (288520838)>,
<ggplot: (288534683)>,
```

```
<ggplot: (278714446)>,
<ggplot: (-9223372036566253168)>,
<ggplot: (-9223372036575101296)>,
<ggplot: (-9223372036566157017)>,
<ggplot: (-9223372036566860821)>,
<ggplot: (286911415)>,
<ggplot: (-9223372036575148270)>,
<ggplot: (-9223372036575153222)>,
<ggplot: (-9223372036575151335)>,
<ggplot: (279627731)>,
<ggplot: (-9223372036575151132)>,
<ggplot: (-9223372036566873933)>,
<ggplot: (-9223372036575151230)>,
<ggplot: (-9223372036566869318)>,
<ggplot: (-9223372036576072767)>,
<ggplot: (-9223372036566863321)>]
```

In [ ]: