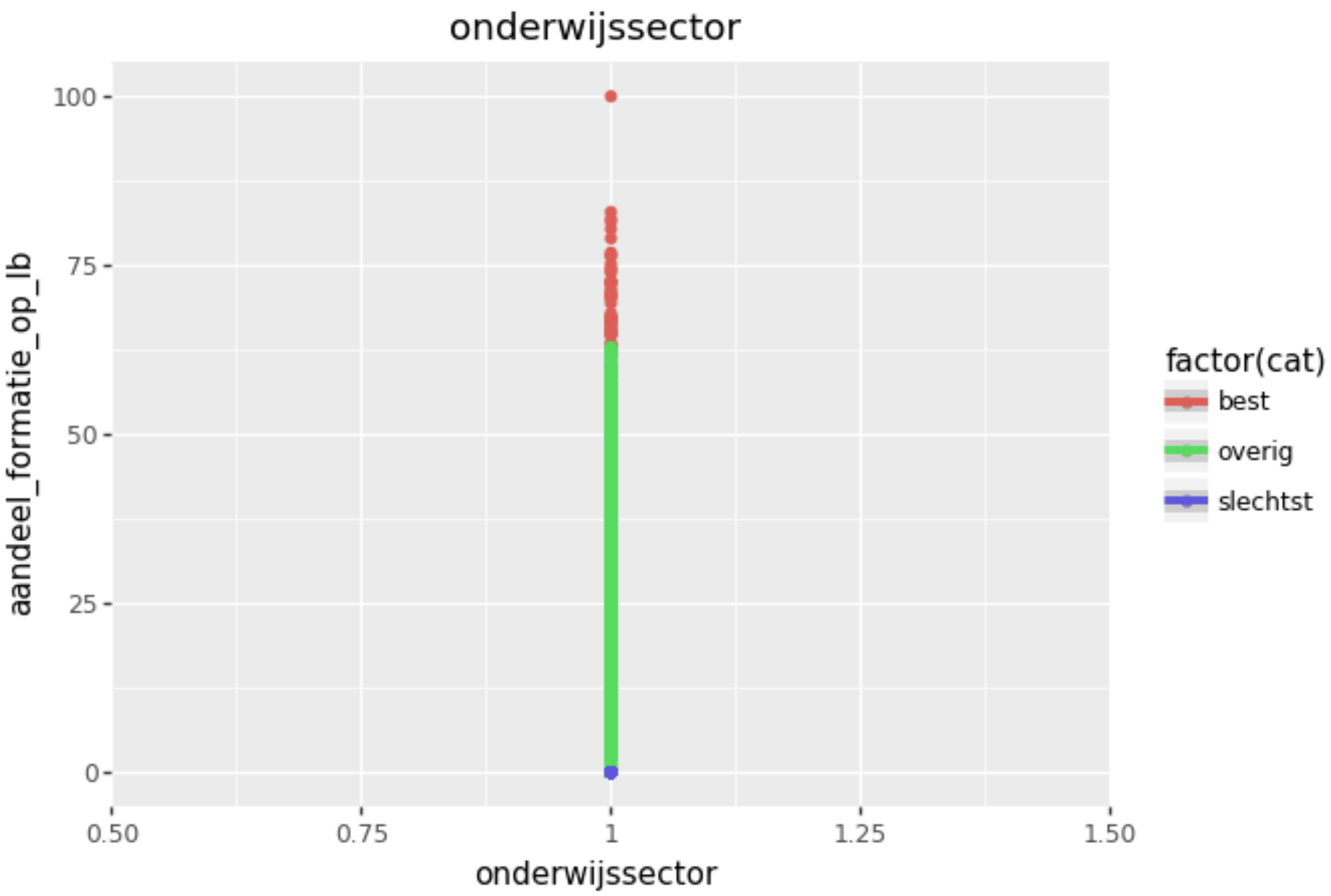
In [1]:

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

root = '/Users/rix0rrr/Google Drive/Hackathon 28_4'
filename = root + '/Team 3/Instellingsdata merged (Elmar)-no-blanks.csv'

#df = pandas.read_csv(filename, index_col='BEVOEGD_GEZAGNAAM', encoding='utf-8')
df = pandas.read_csv(filename, index_col='key', encoding='utf-8', low_memory=Fals
    'aandeel_formatie_op_lb': np.float64
})

# 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['year'] == 2016]

# geen kleine scholen
df = df.loc[df['omvang_formatie_totaal'] >= 5]

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

print("Instellingen 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_GEZAG
#df = df[:-1]
#cols = [c for c in cols if not c.startswith('AANDEEL_FORMATIE_')]

graphs = []

cols = [c for c in cols if c not in ['BRIN_NUMMER', 'instellingsnaam', 'bevoegd g
```

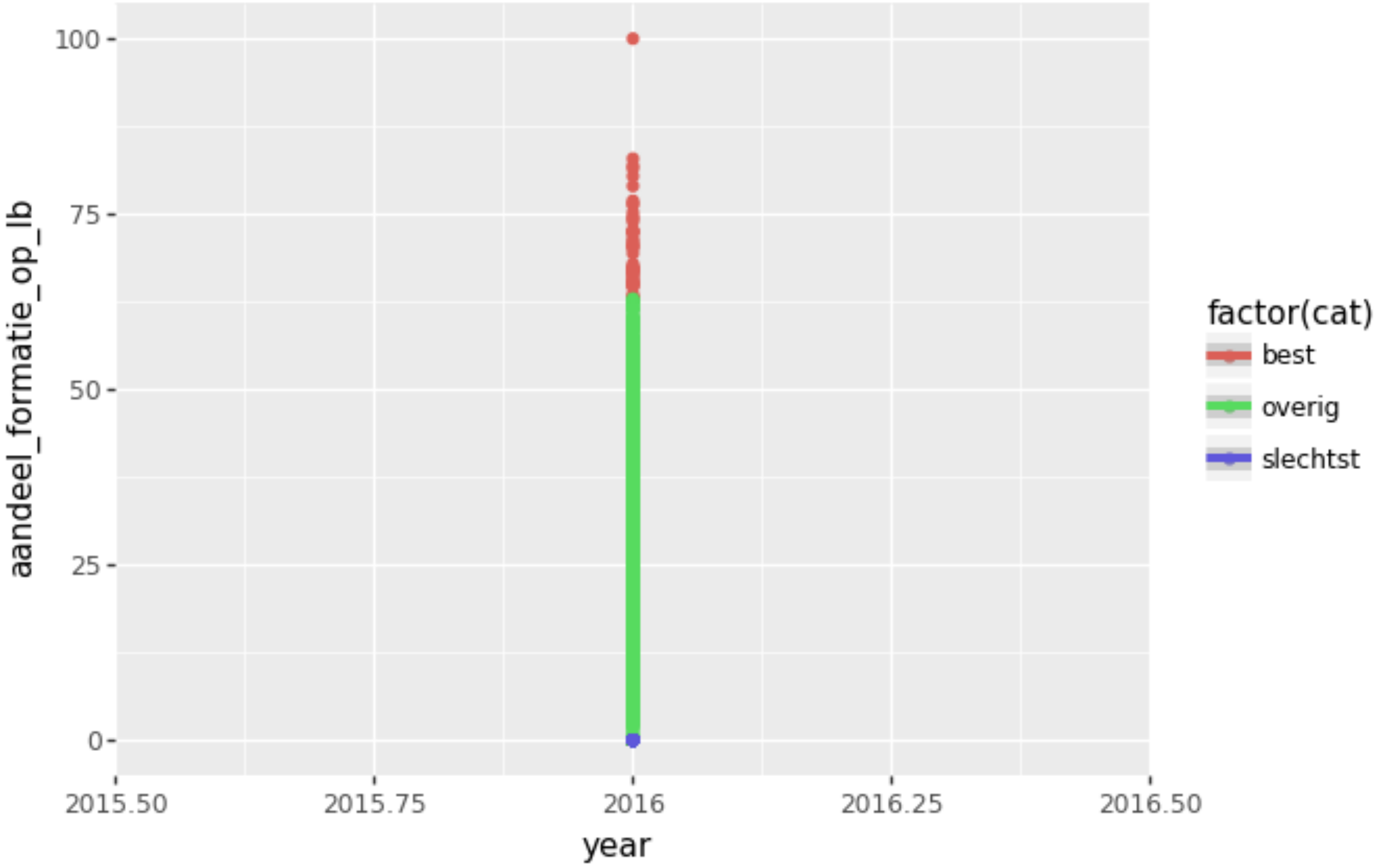
```
for x_as in cols:
    print(x_as)
    #x_as = 'SolvabiliteitI.2016'
    y_as = 'aandeel_formatie_op_lb'
    try:
        rows = df[[x_as, y_as, 'cat']].dropna()
        if not len(rows):
            print('NO NON-NA ROWS')
            continue

        print(ggplot(rows, aes(x_as, y_as, color='factor(cat)'))
         + ggtitle(x_as)
         + geom_point()
         + stat_smooth(method='lm'))
    except Exception as e:
        traceback.print_exc()
        print(e)
        pass

graphs
```

```
/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/site-pac
kages/statsmodels/compat/pandas.py:56: FutureWarning: The pandas.cor
e.datetools module is deprecated and will be removed in a future ver
sion. Please use the pandas.tseries module instead.
  from pandas.core import datetools

Instellingen in analyse: 5902
onderwijssector
```

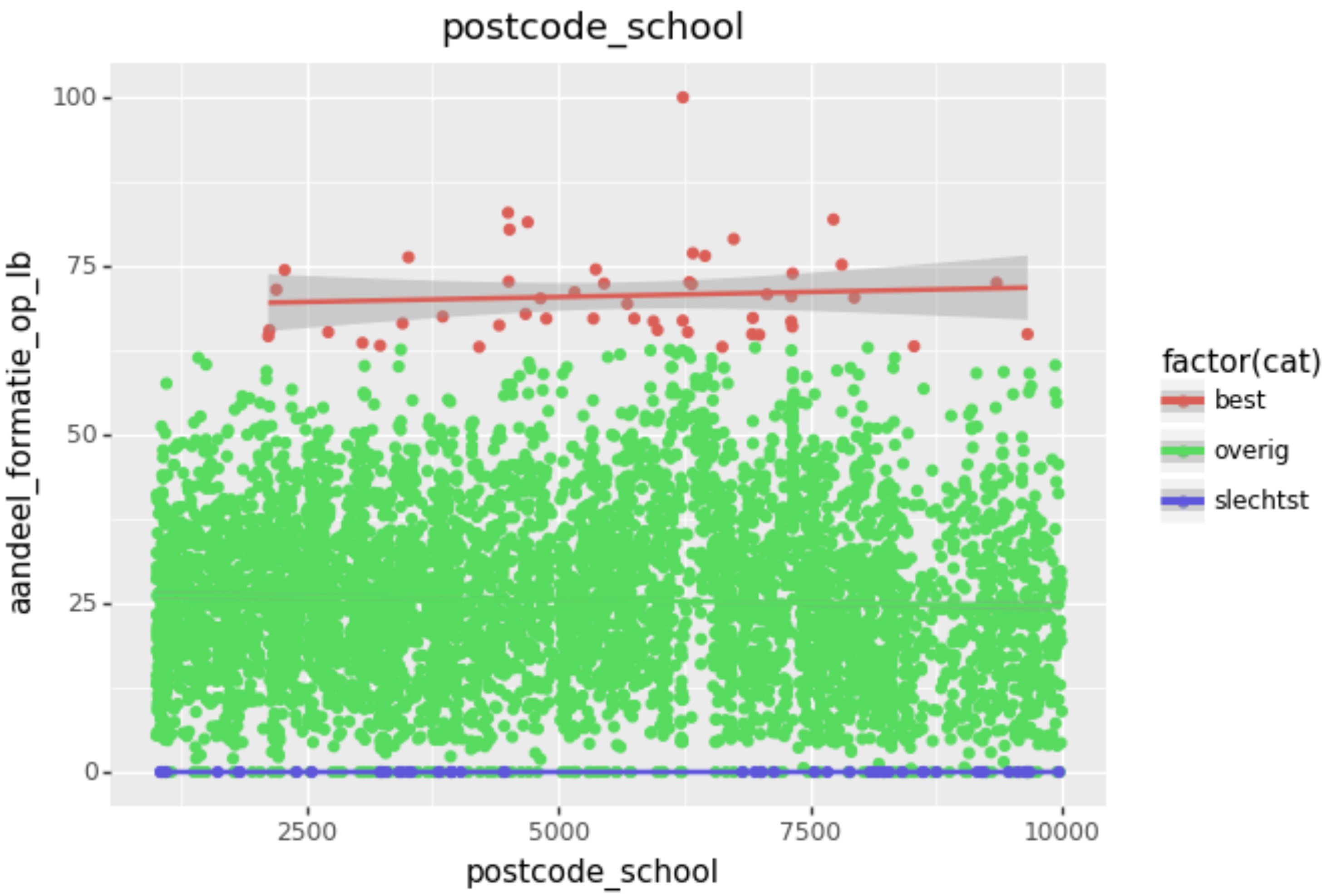


```
<ggplot: (287230514)>
year
```

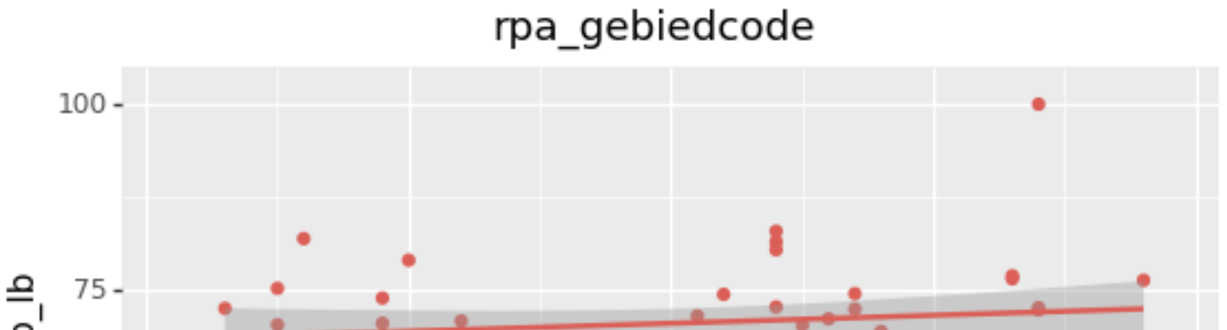
year

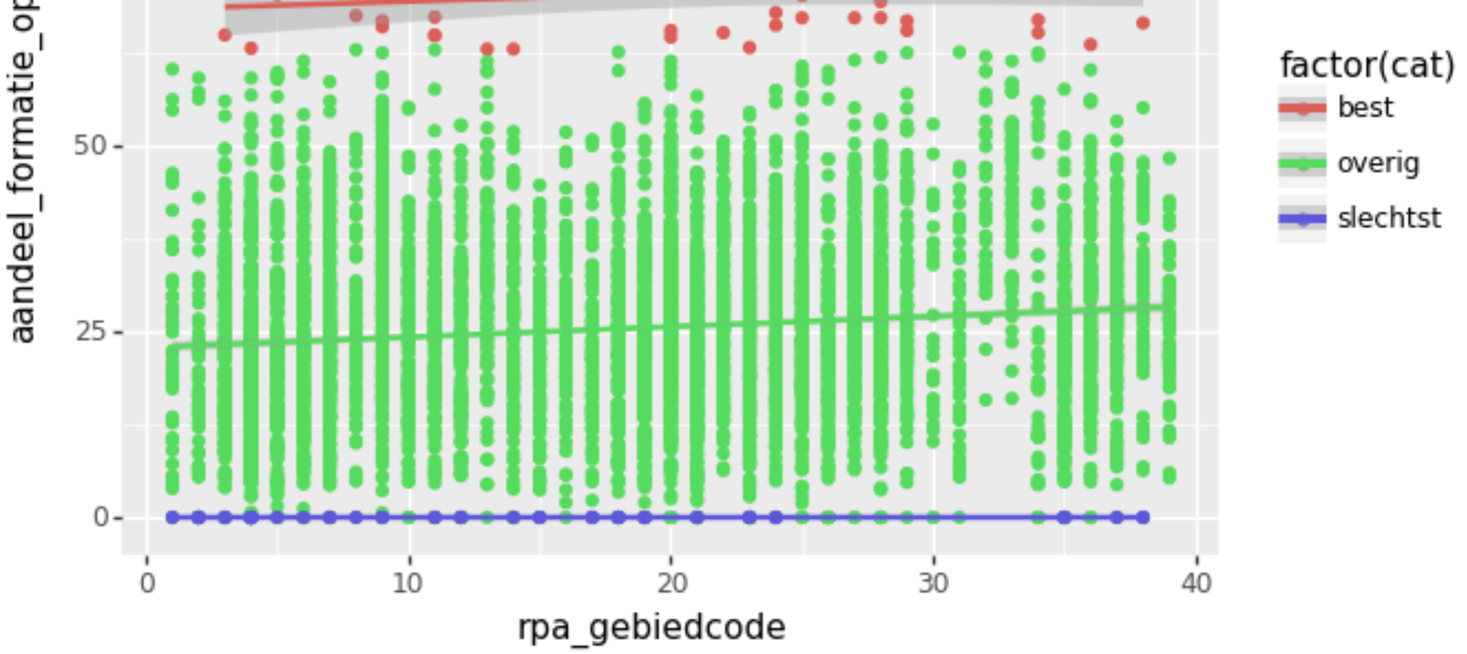


```
<ggplot: (-9223372036568461133)>
postcode_school
```

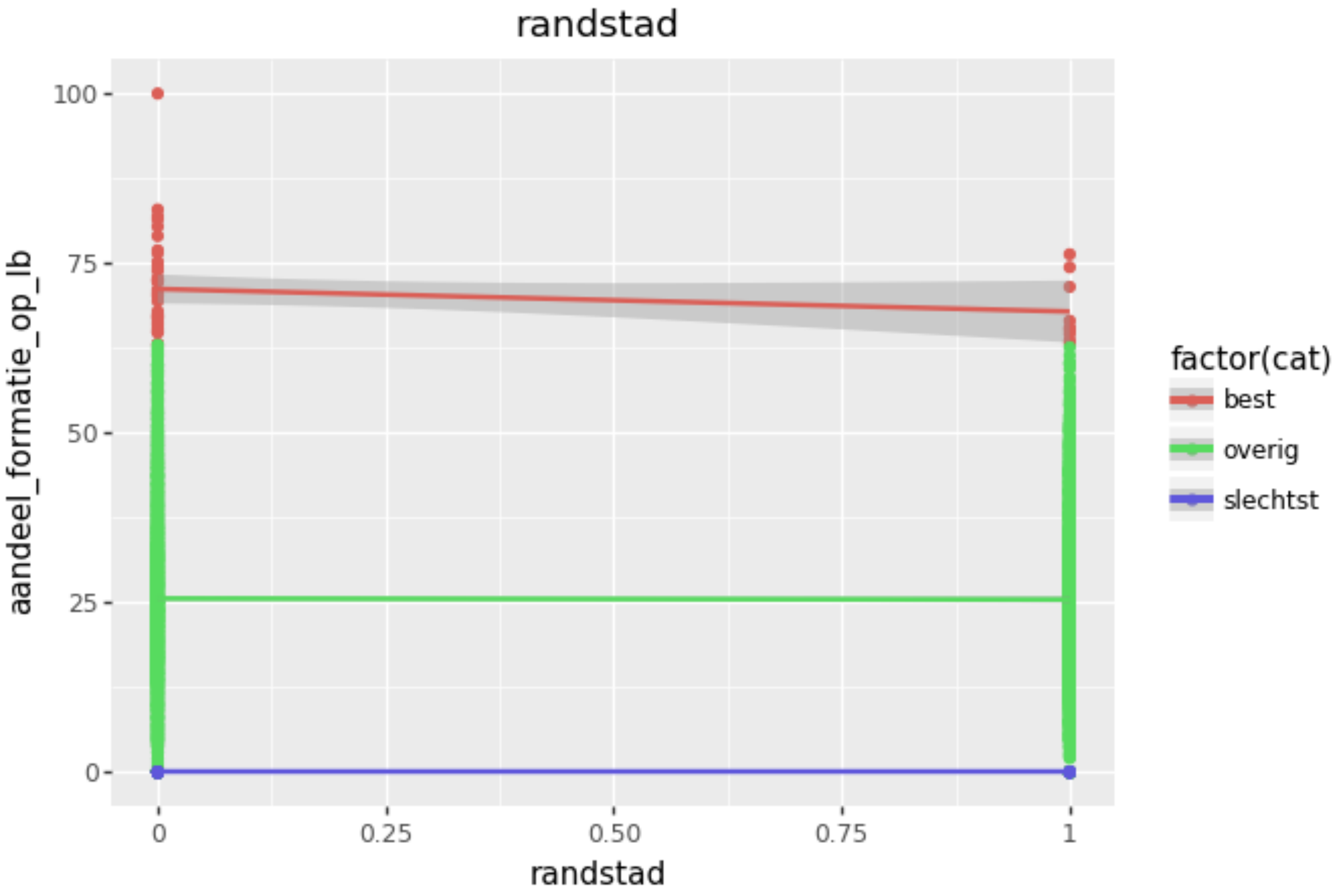


```
<ggplot: (-9223372036569254119)>
rpa_gebiedcode
```

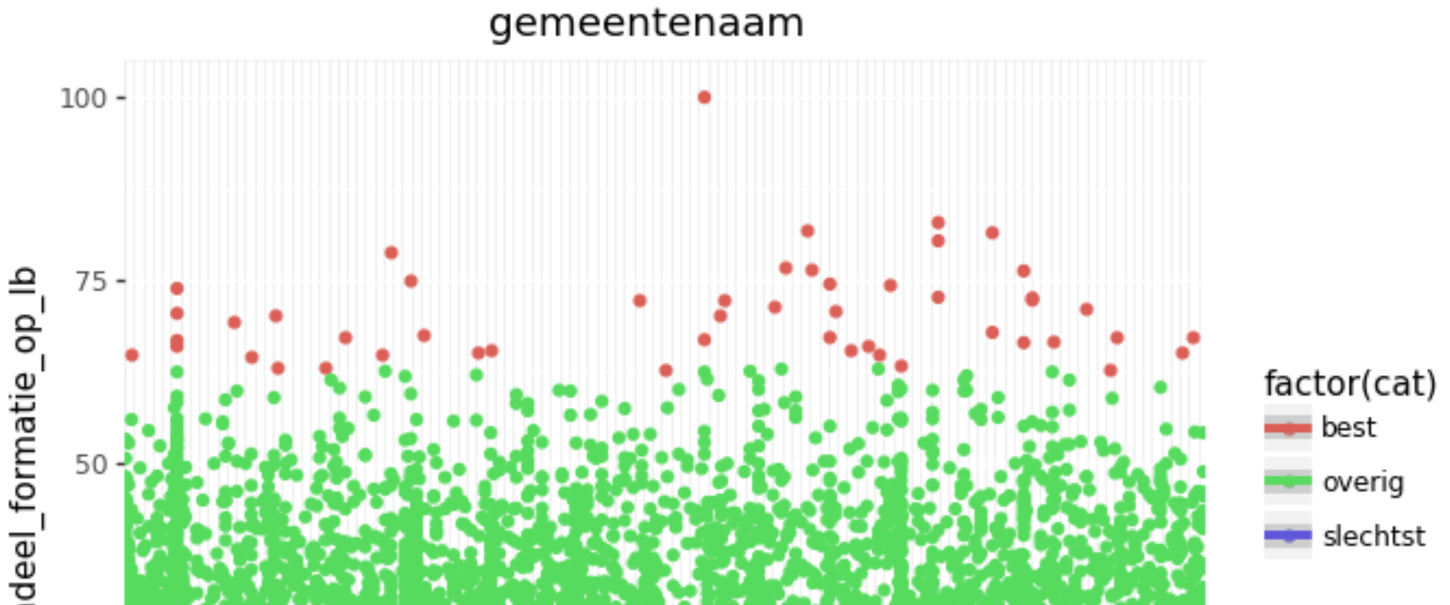


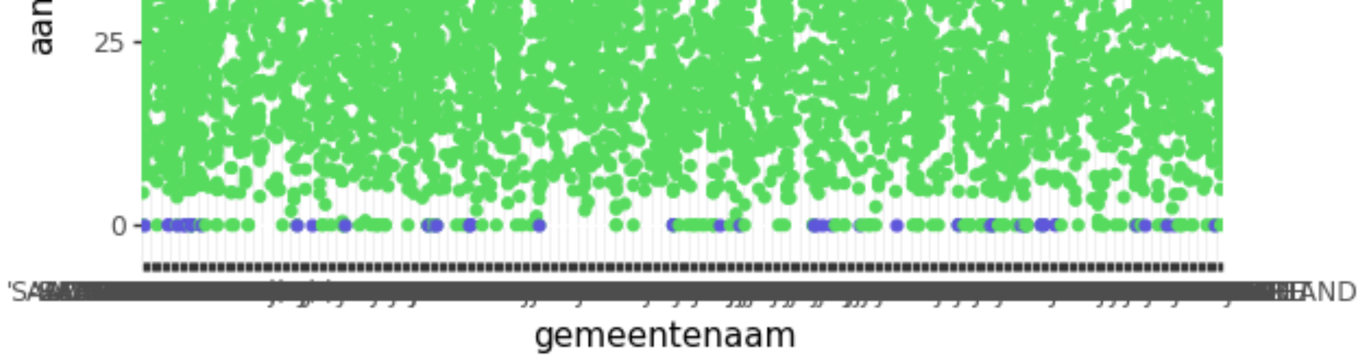


```
<ggplot: (287225306)>
randstad
```

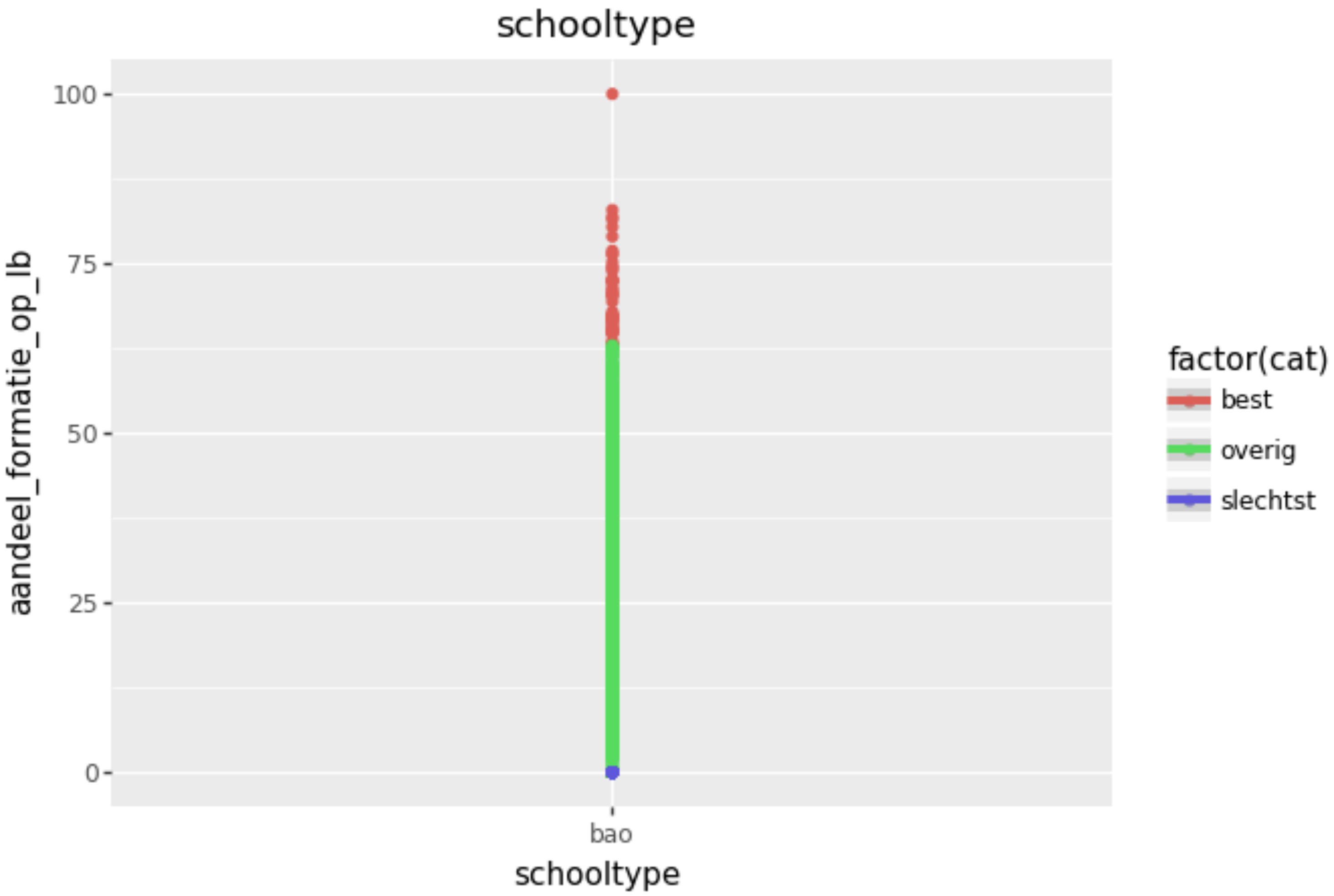


```
<ggplot: (285878462)>
gemeentenaam
```

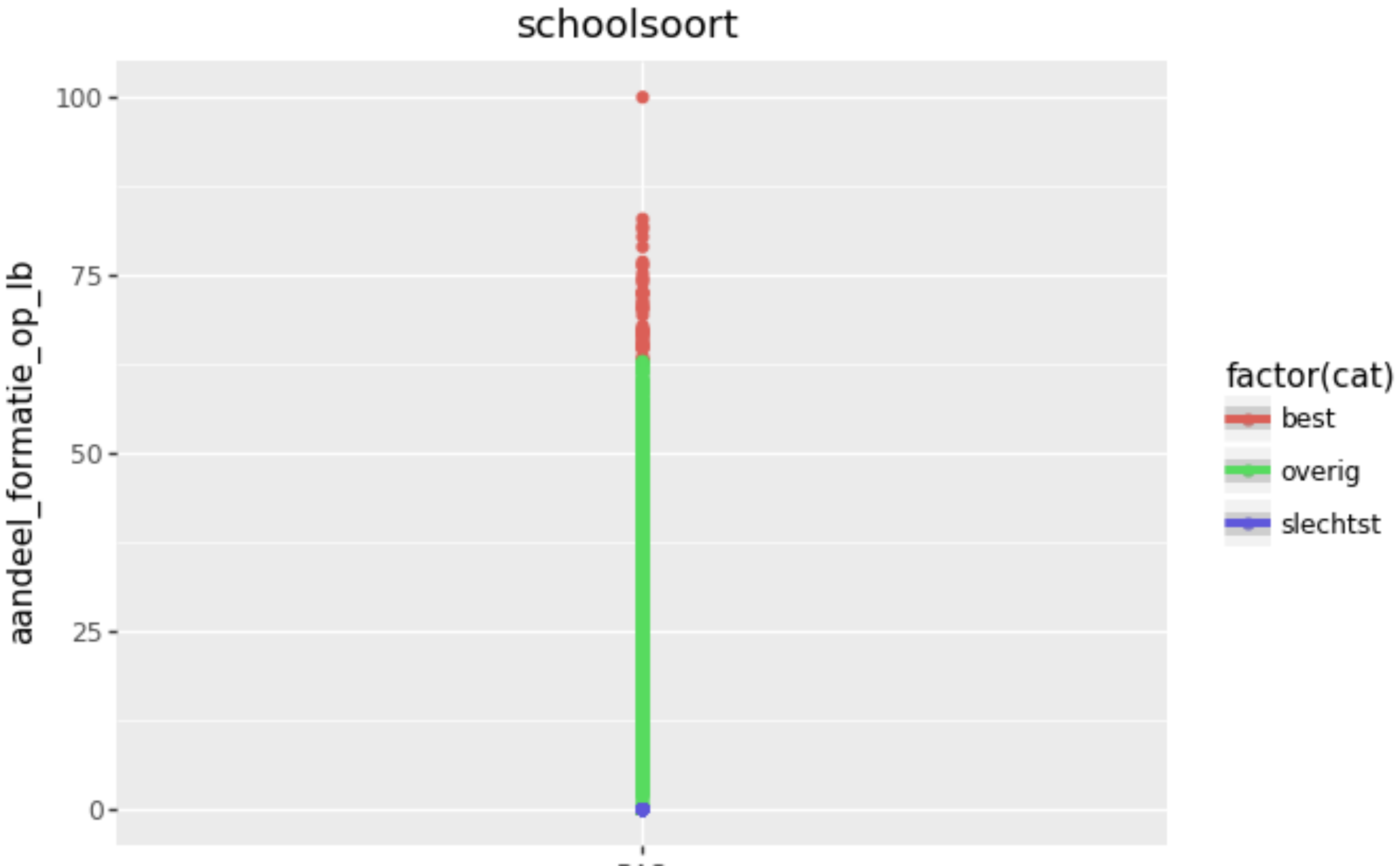


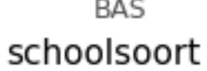


```
<ggplot: (-9223372036567486534)>
schooltype
```

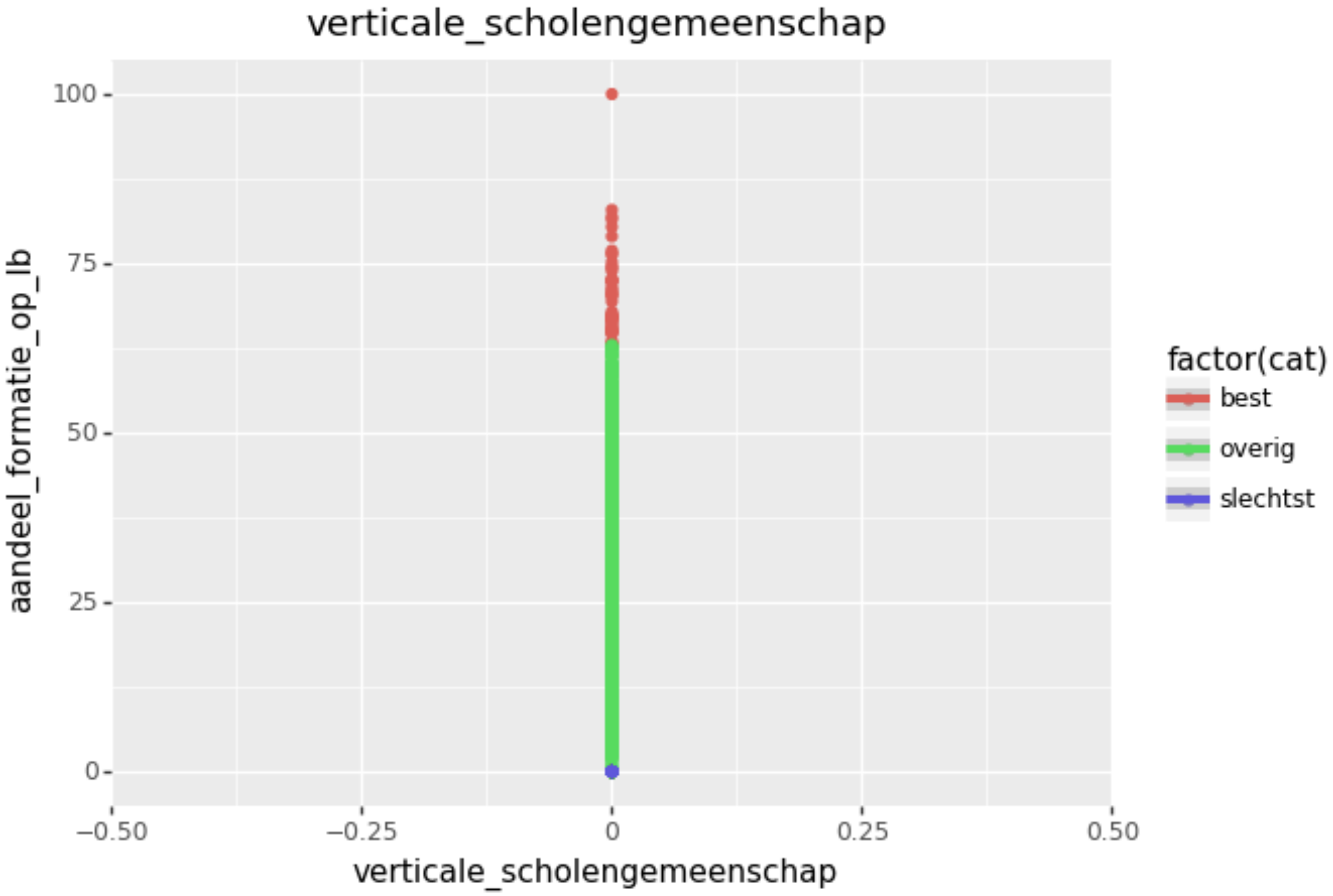


```
<ggplot: (-9223372036569251952)>
schoolsoort
```

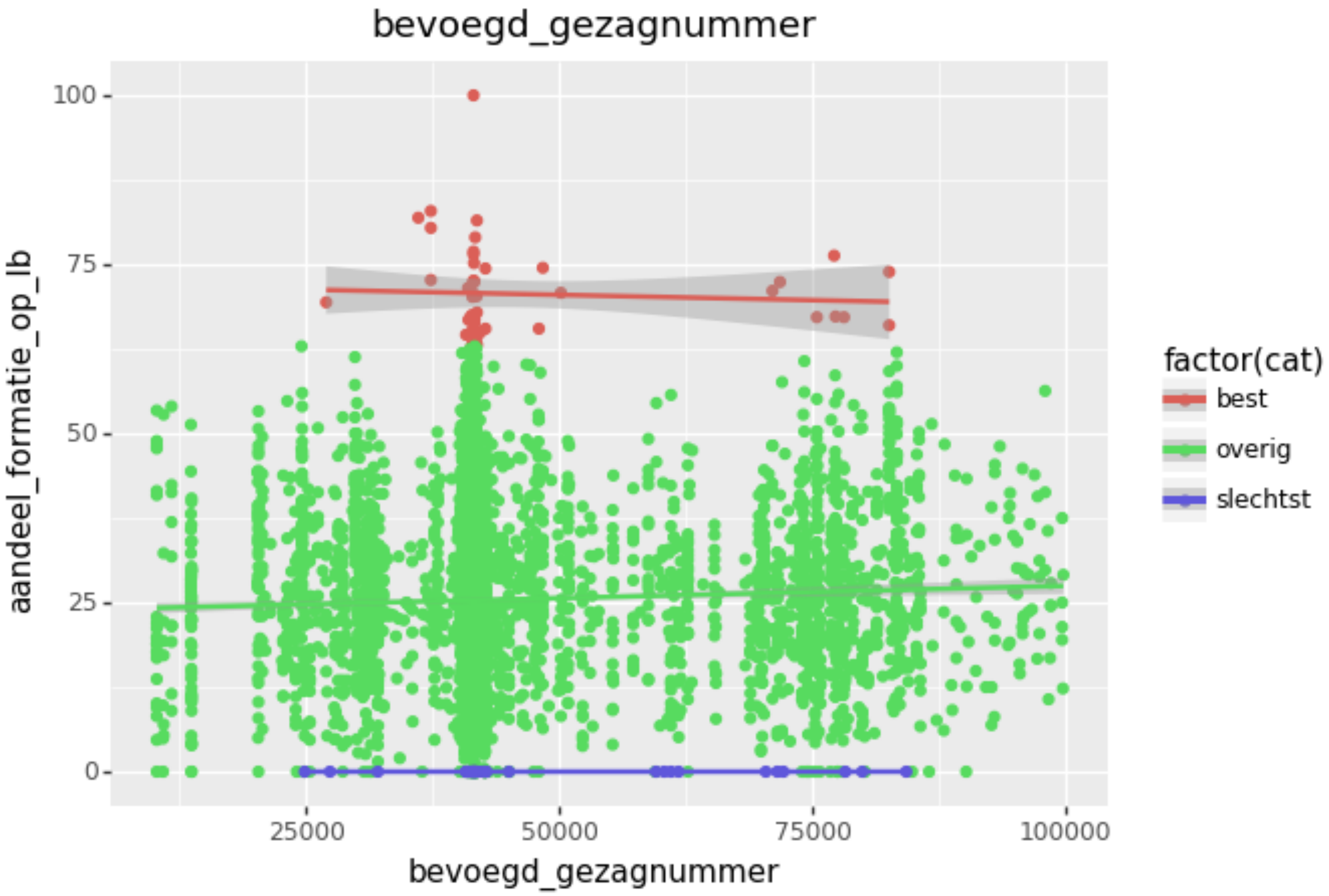




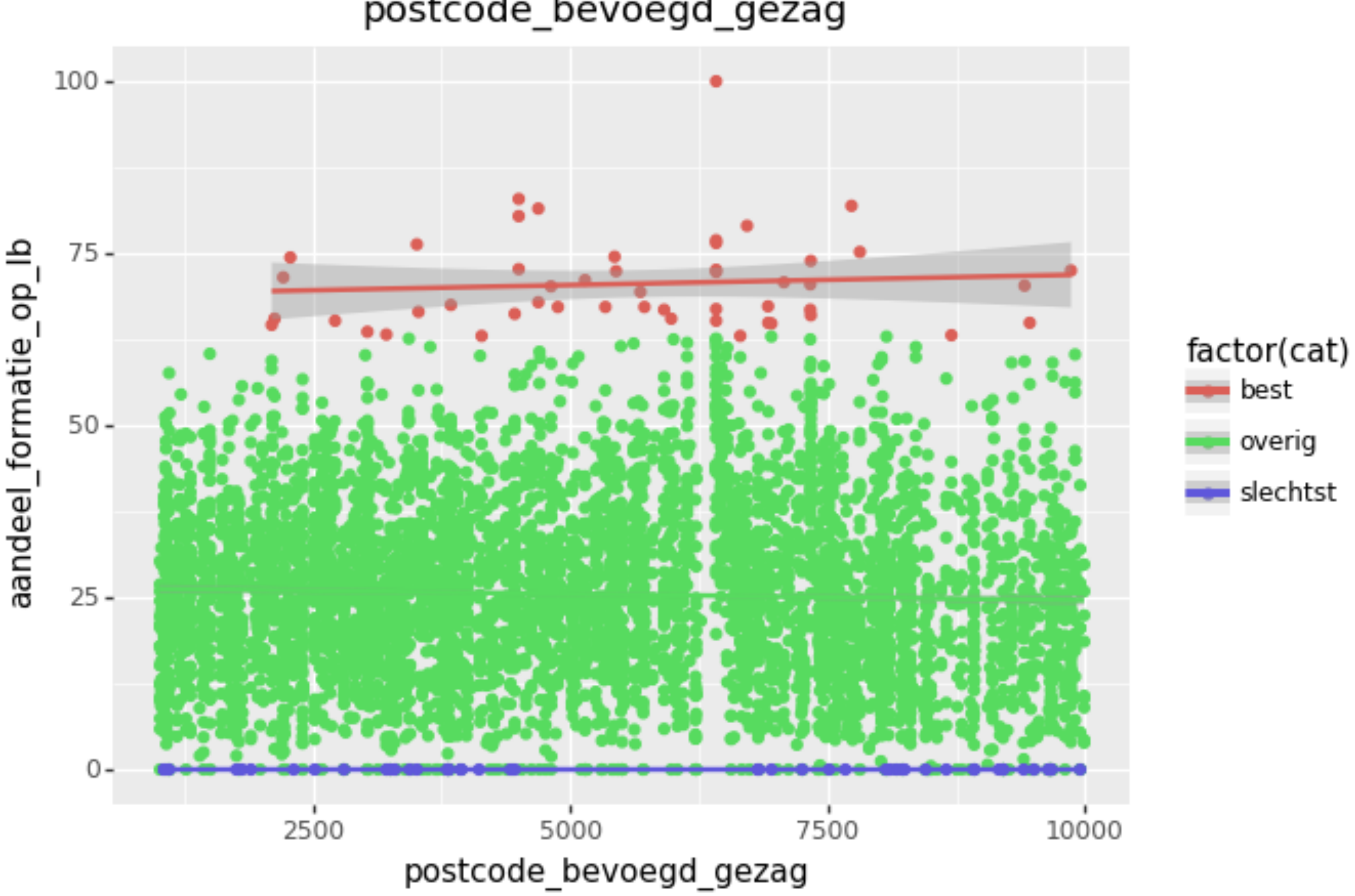
```
<ggplot: (286747940)>
verticale_scholengemeenschap
```



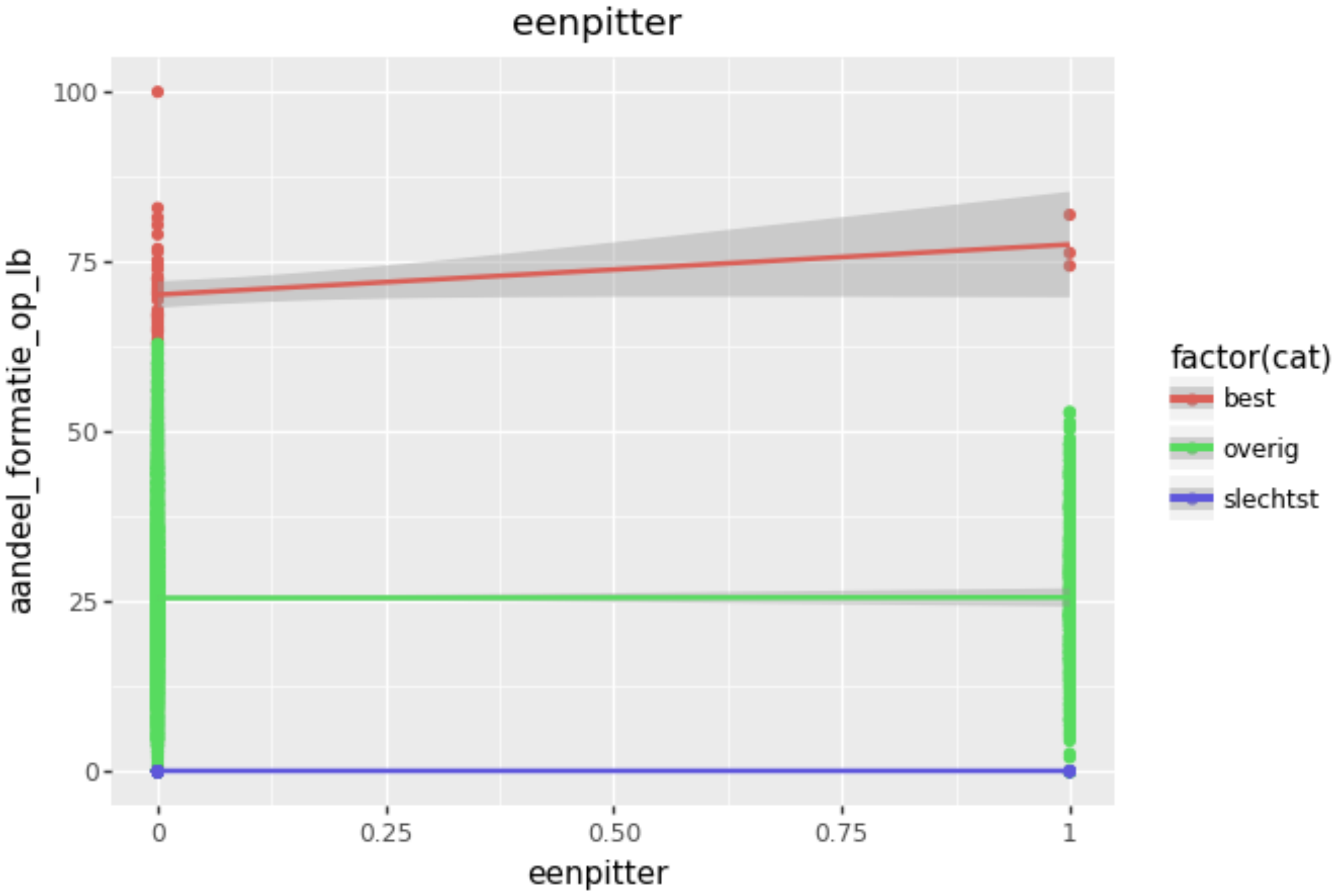
```
<ggplot: (-9223372036565884500)>
bevoegd_gezagnummer
```



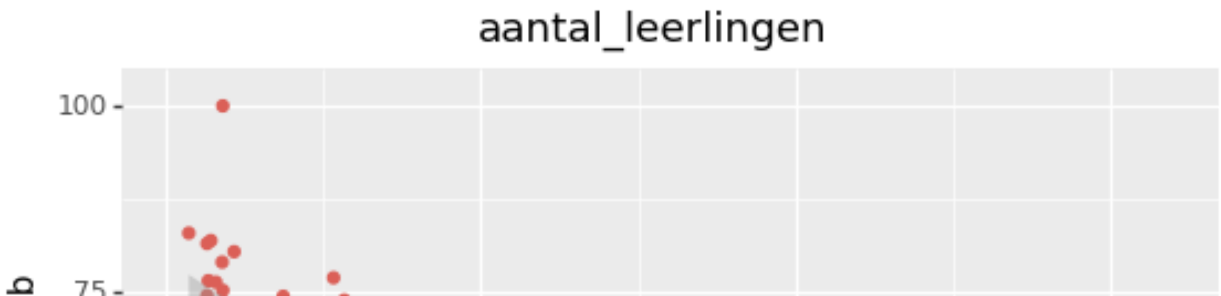
```
<ggplot: (285523709)>
postcode_bevoegd_gezag
```

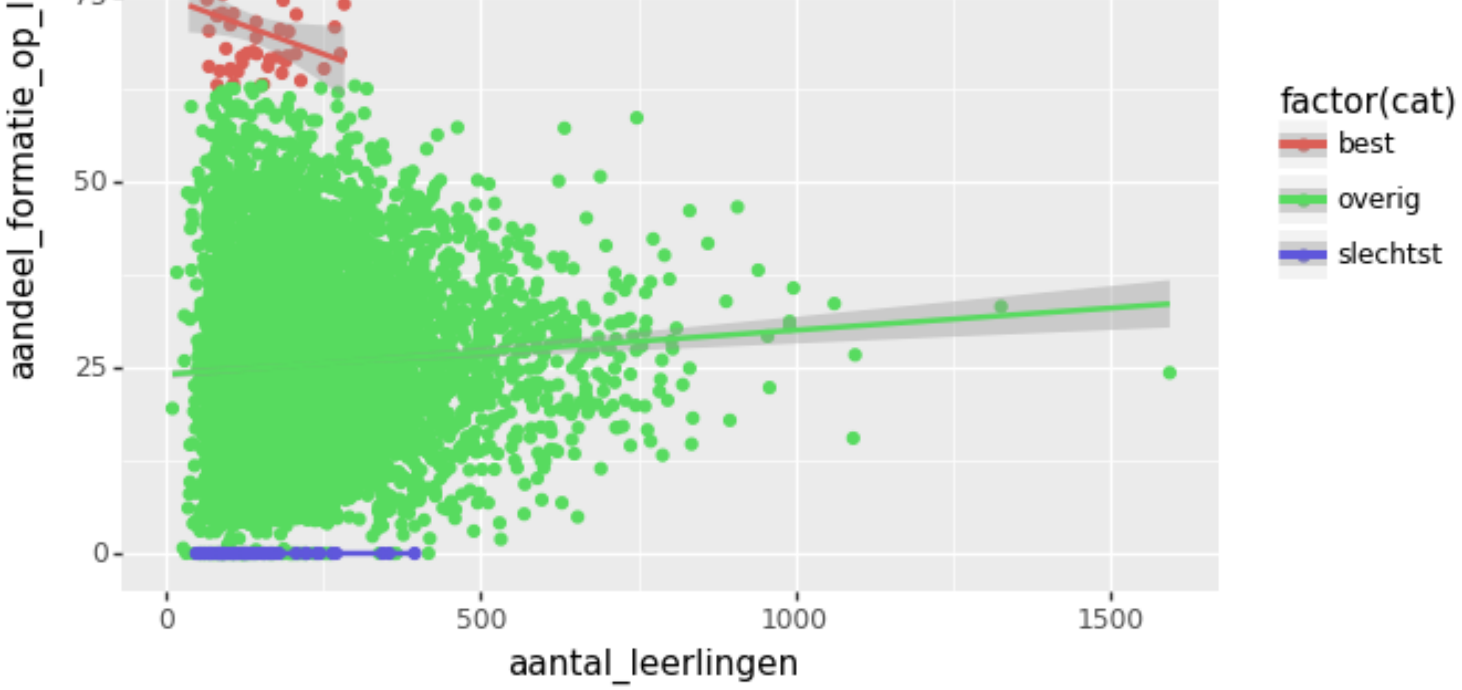


```
<ggplot: (289125026)>
eenpitter
```

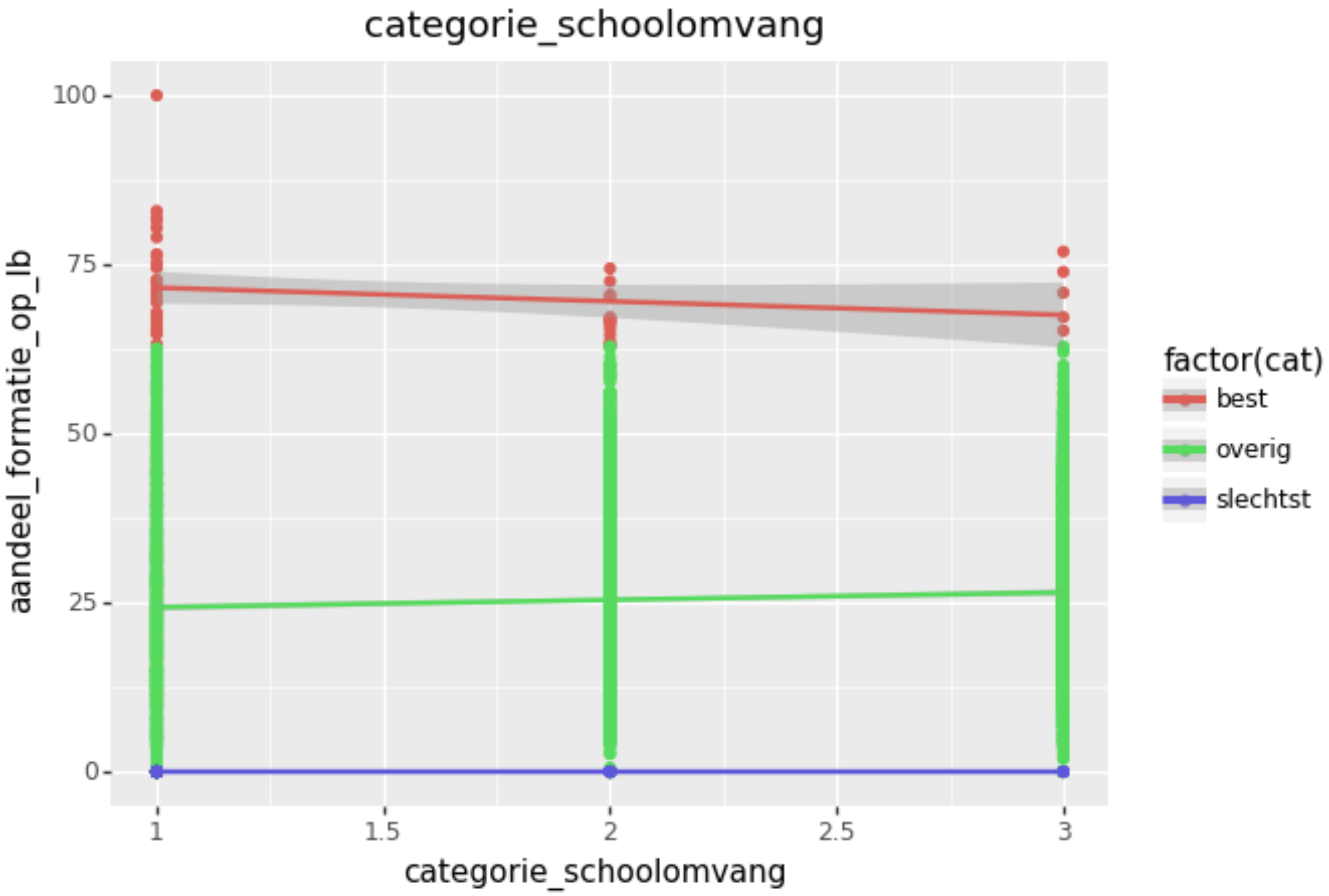


```
<ggplot: (-9223372036565650772)>
aantal_leerlingen
```

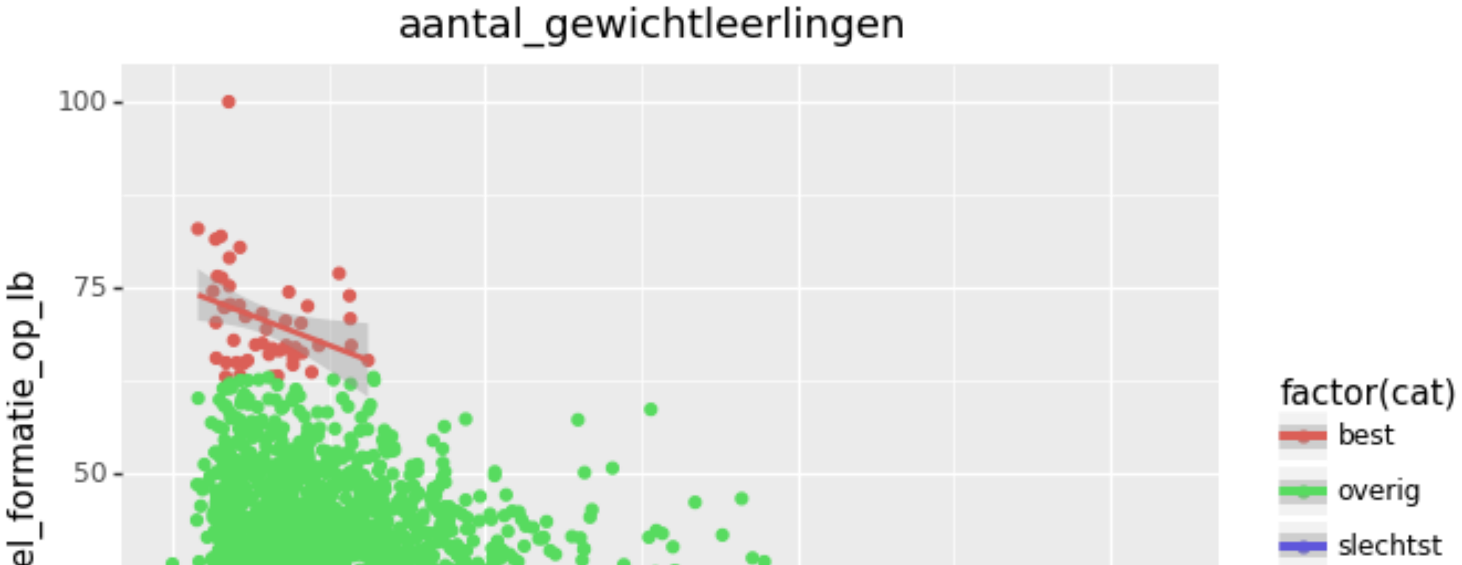


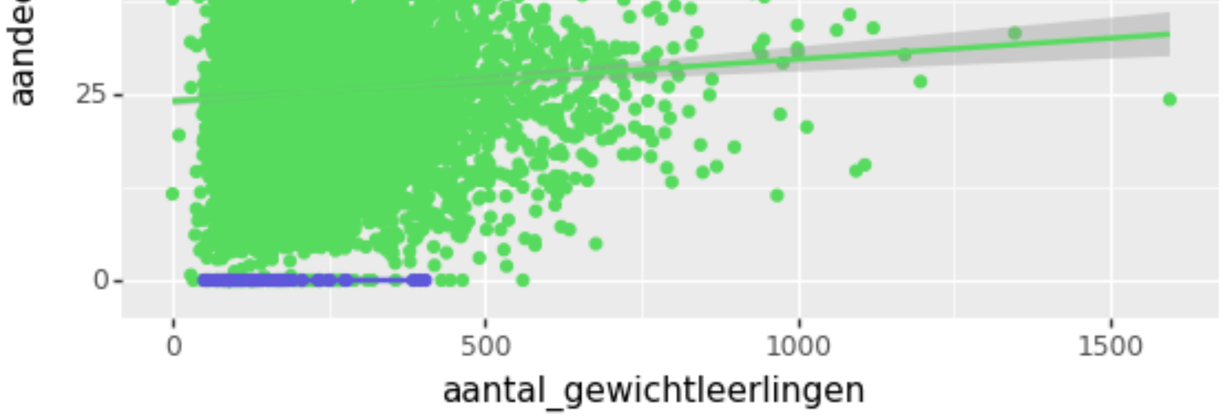


```
<ggplot: (289521928)>
categorie_schoolomvang
```

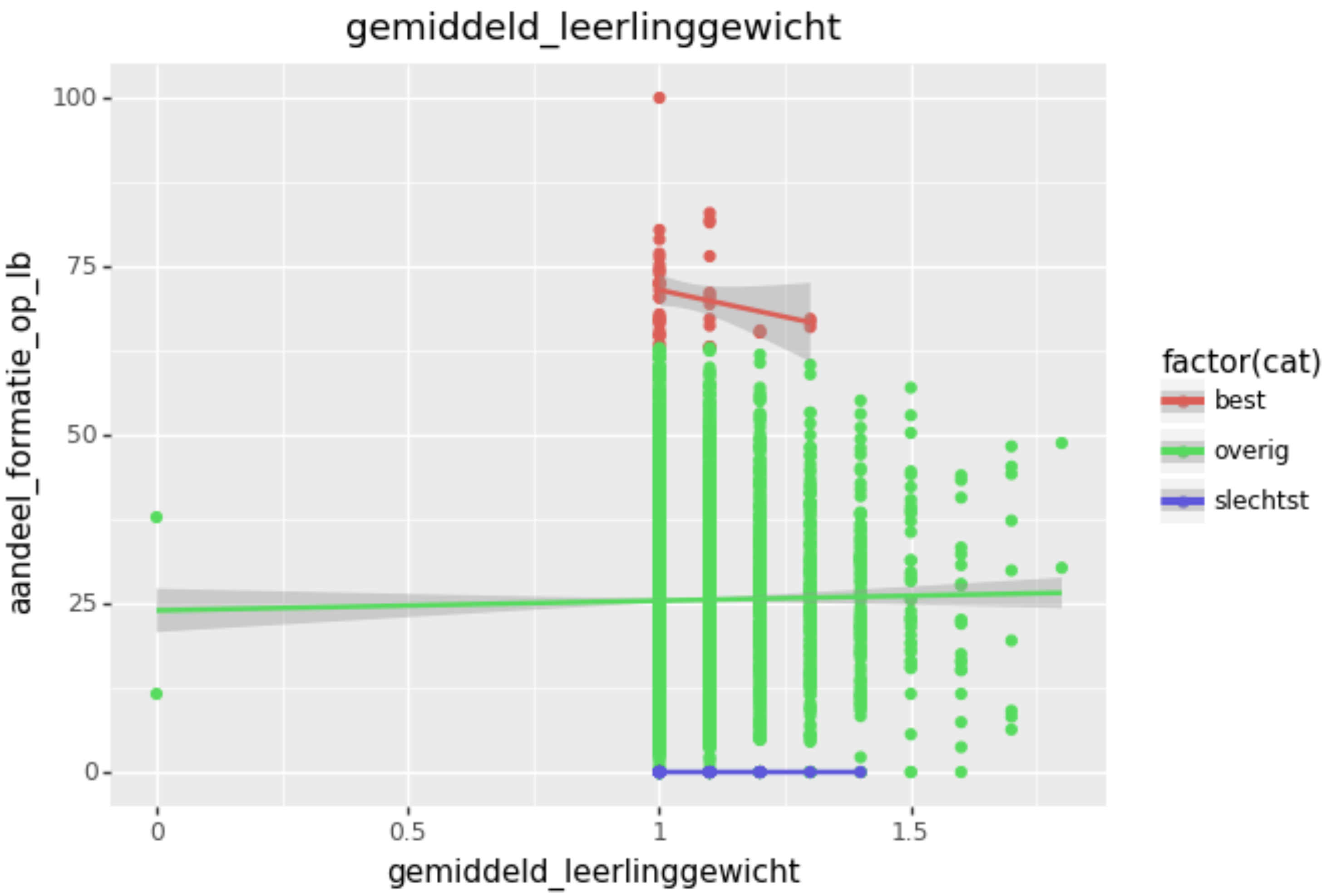


```
<ggplot: (289836762)>
aantal_gewichtleerlingen
```

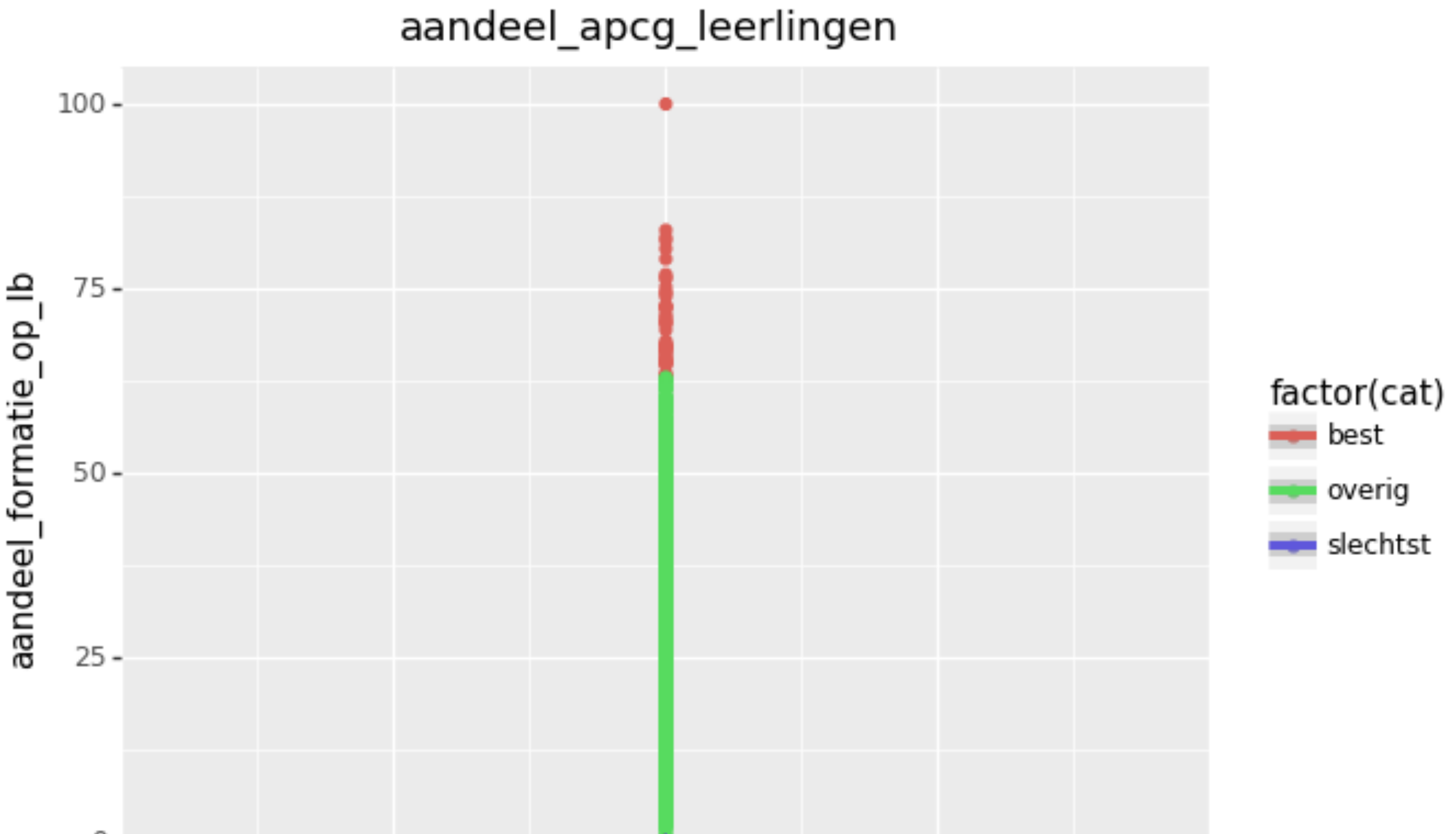


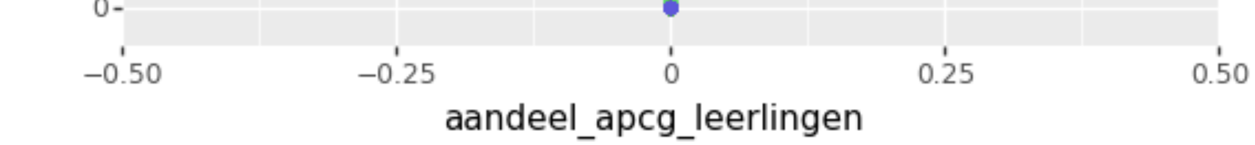


```
<ggplot: (-9223372036568573815)>
gemiddeld_leerlinggewicht
```

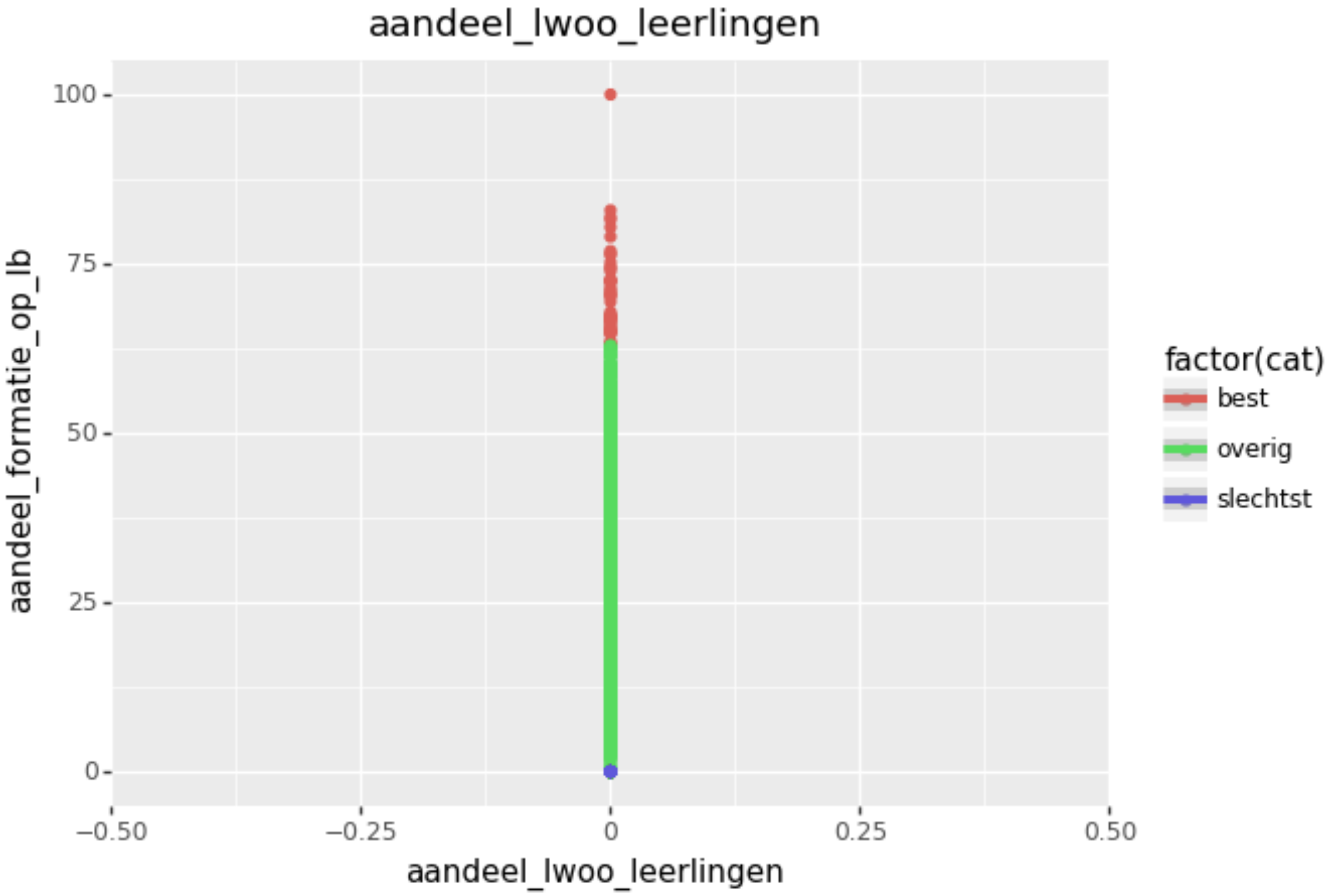


```
<ggplot: (-9223372036564367714)>
aandeel_apcg_leerlingen
```

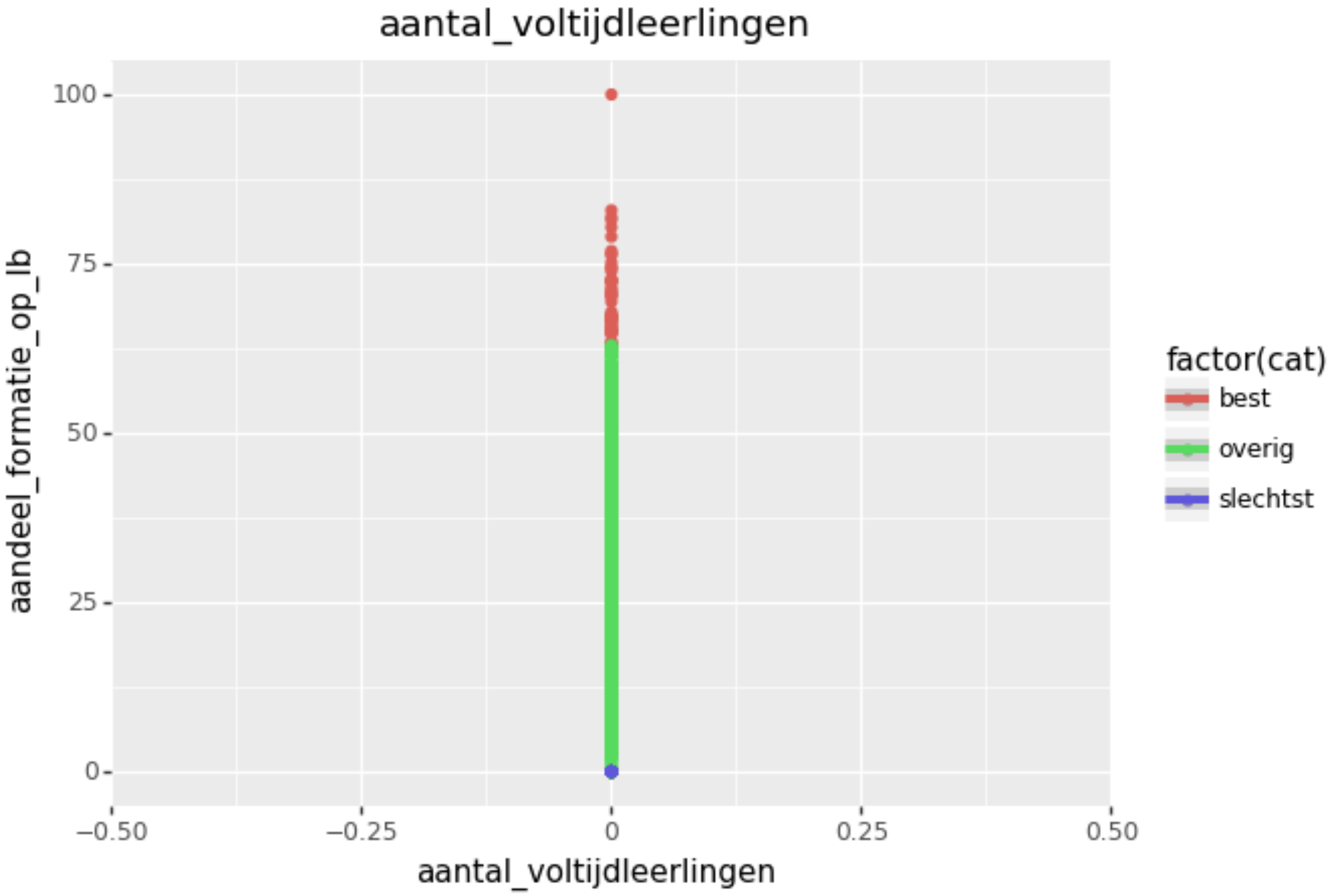




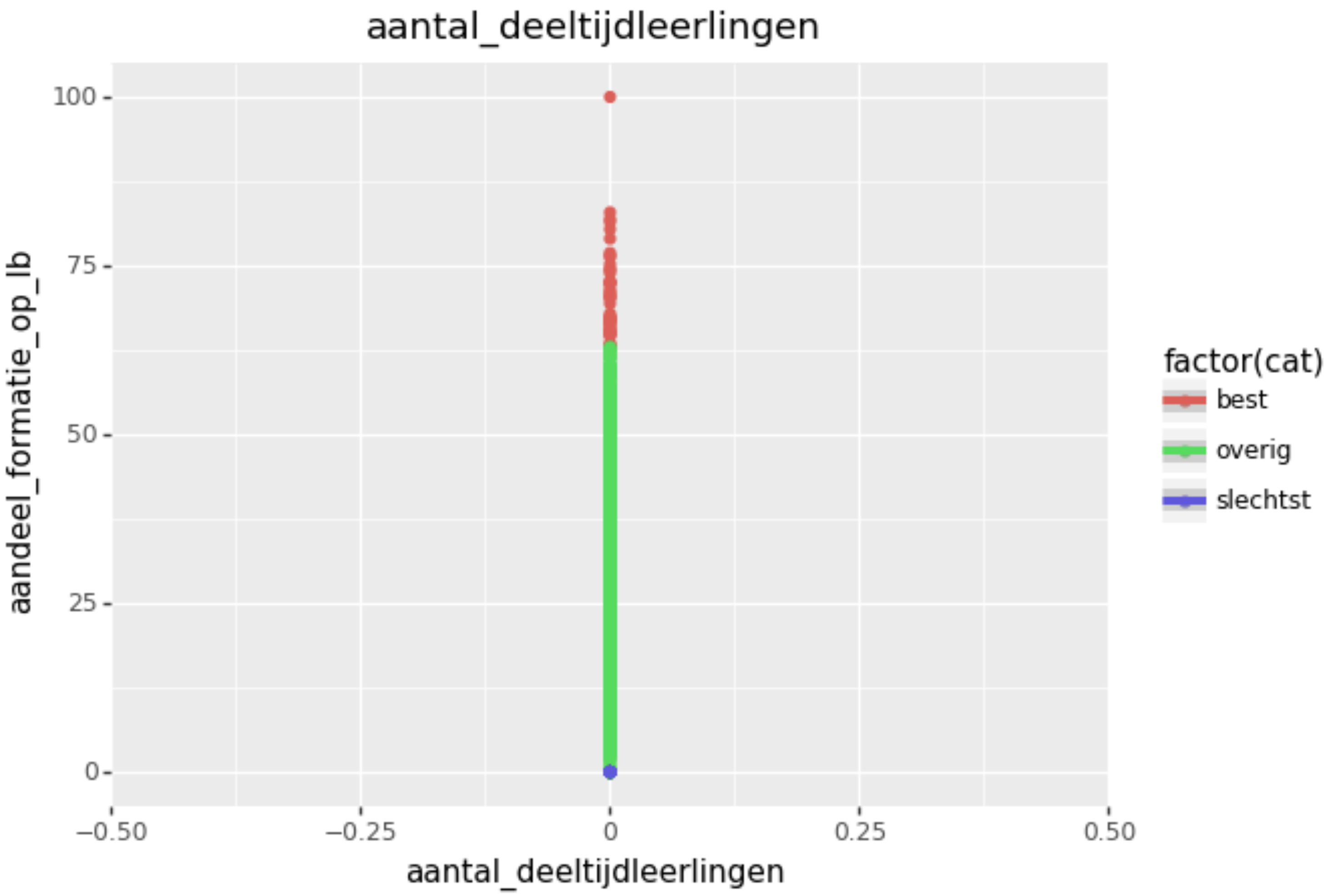
```
<ggplot: (-9223372036564137833)>
aandeel_lwoo_leerlingen
```



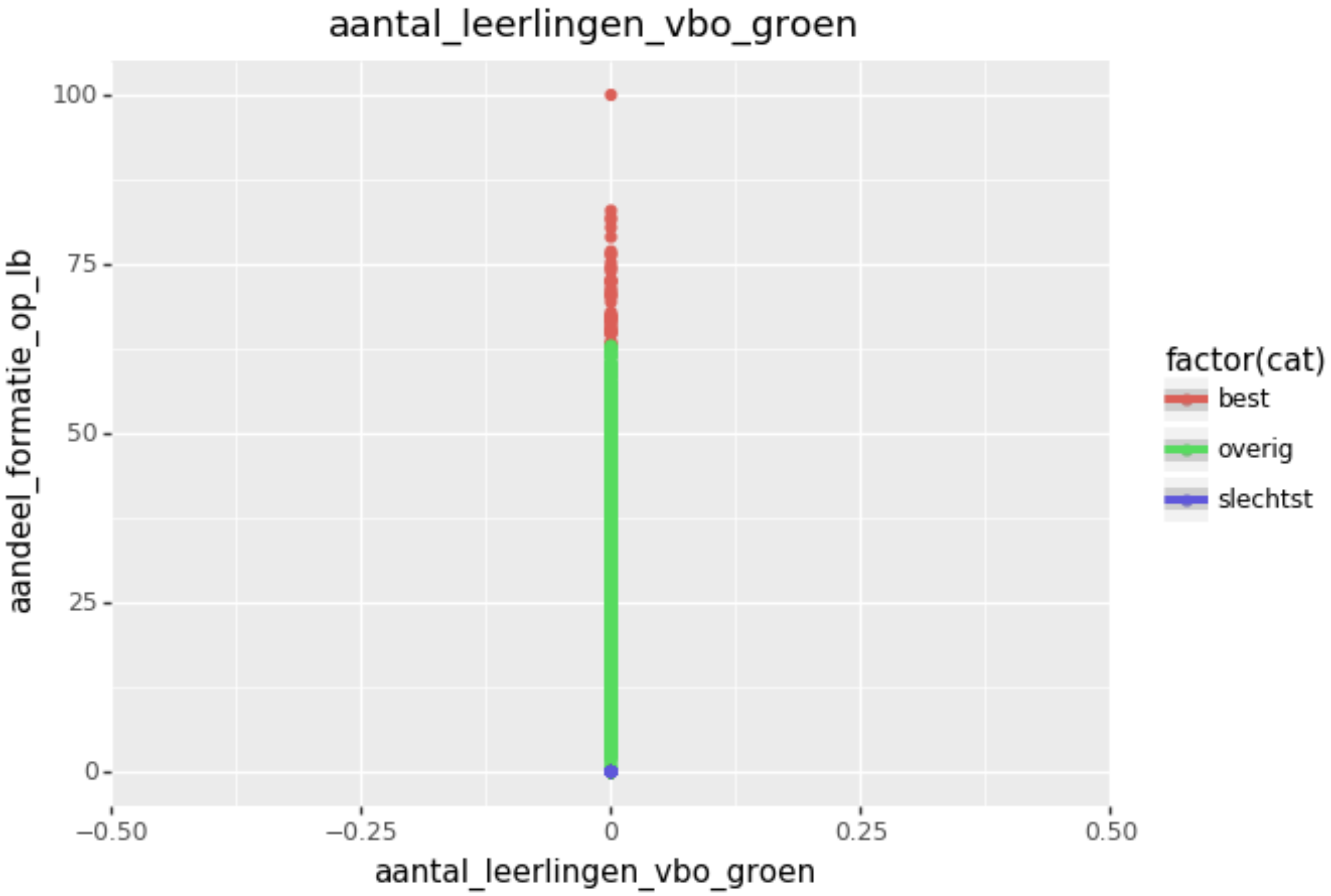
```
<ggplot: (290638063)>
aantal_voltijdleerlingen
```



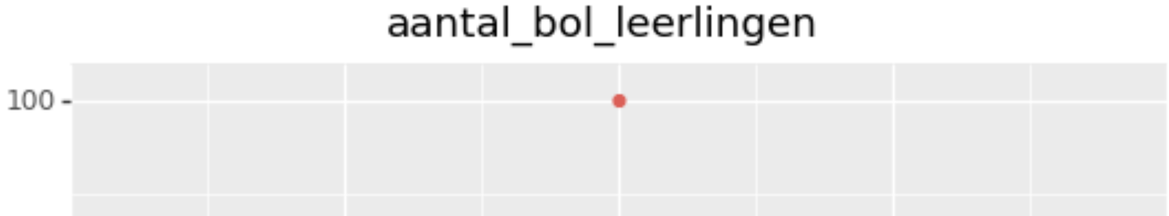
```
<ggplot: (291450095)>
aantal_deeltijdleerlingen
```

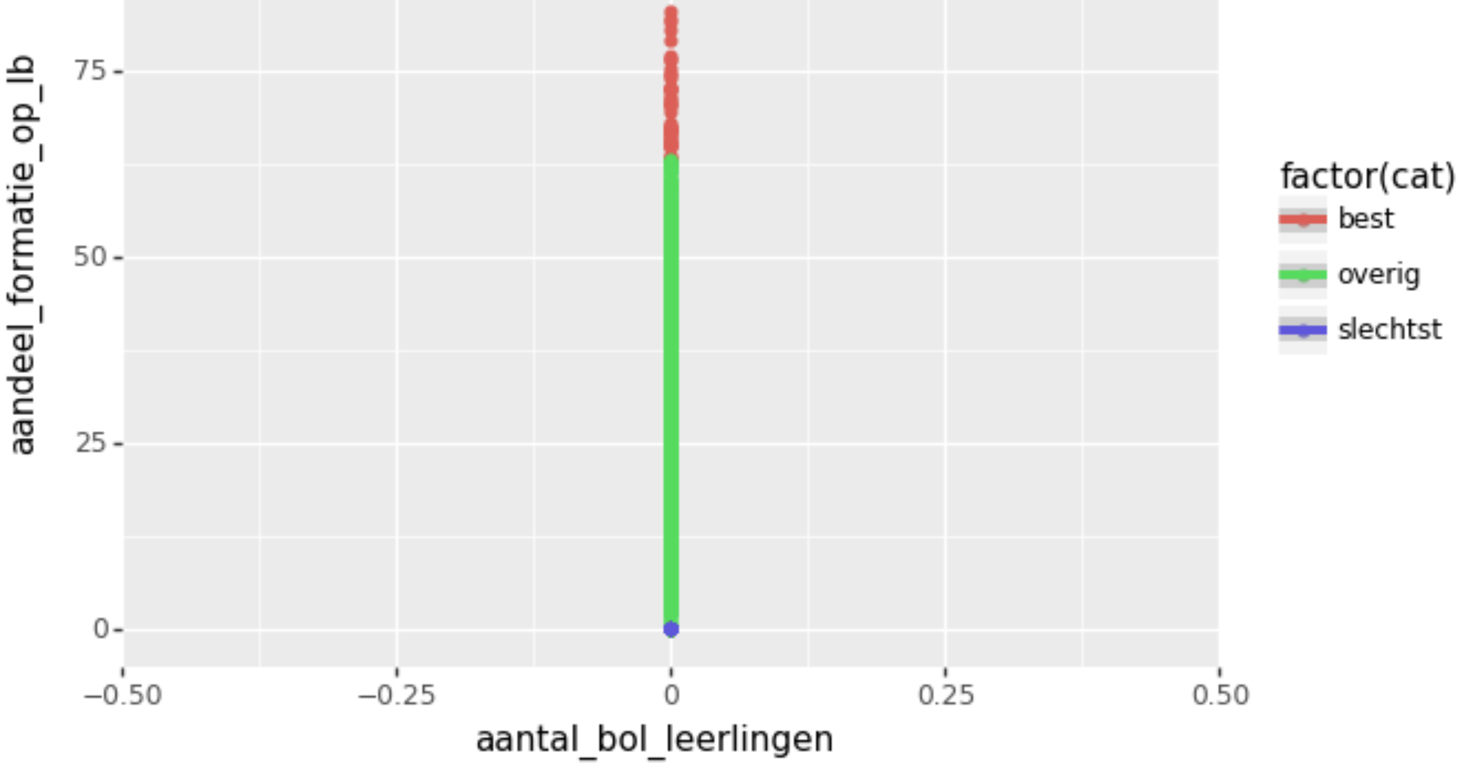


```
<ggplot: (-9223372036562949425)>
aantal_leerlingen_vbo_groen
```

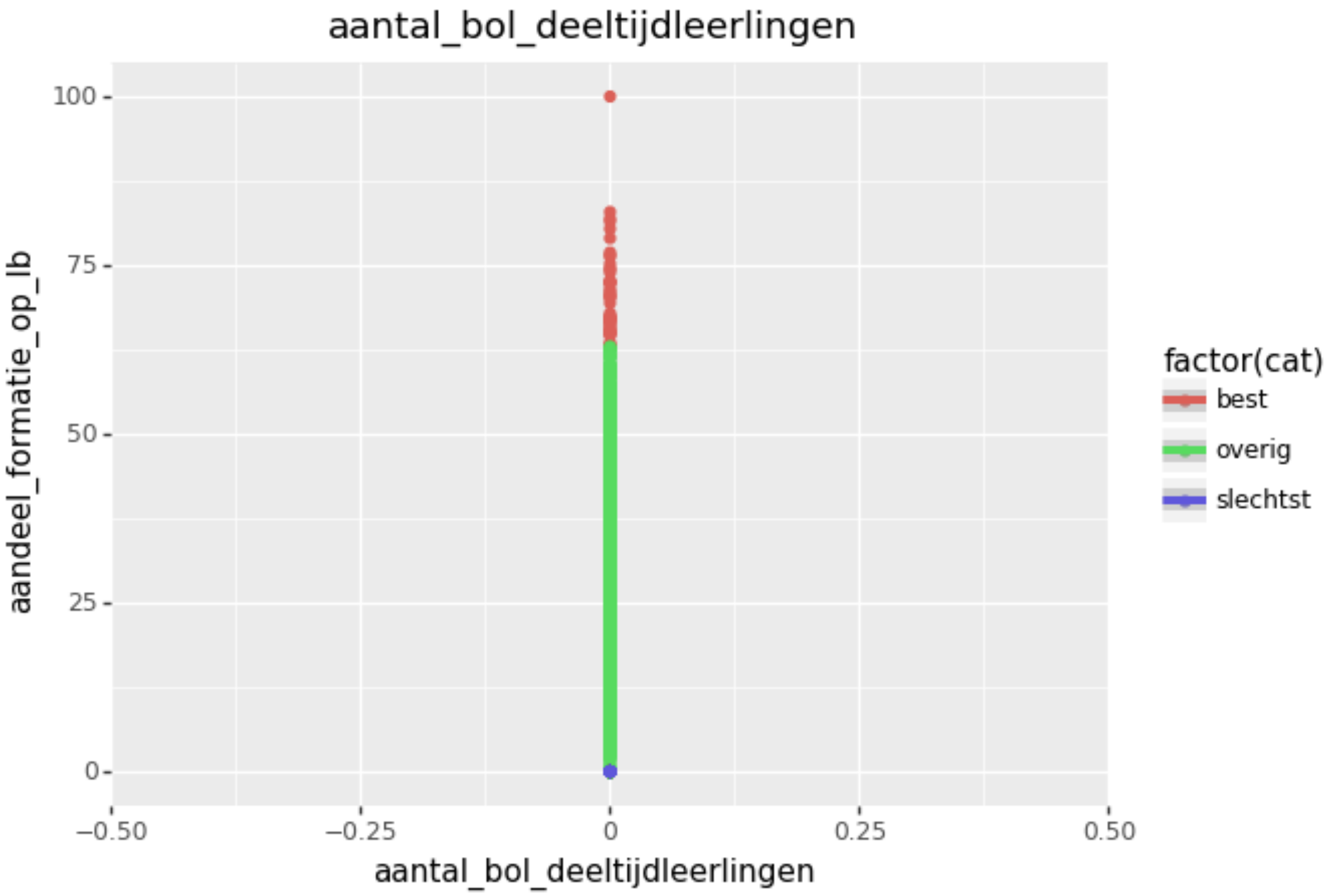


```
<ggplot: (-9223372036562670406)>
aantal_bol_leerlingen
```

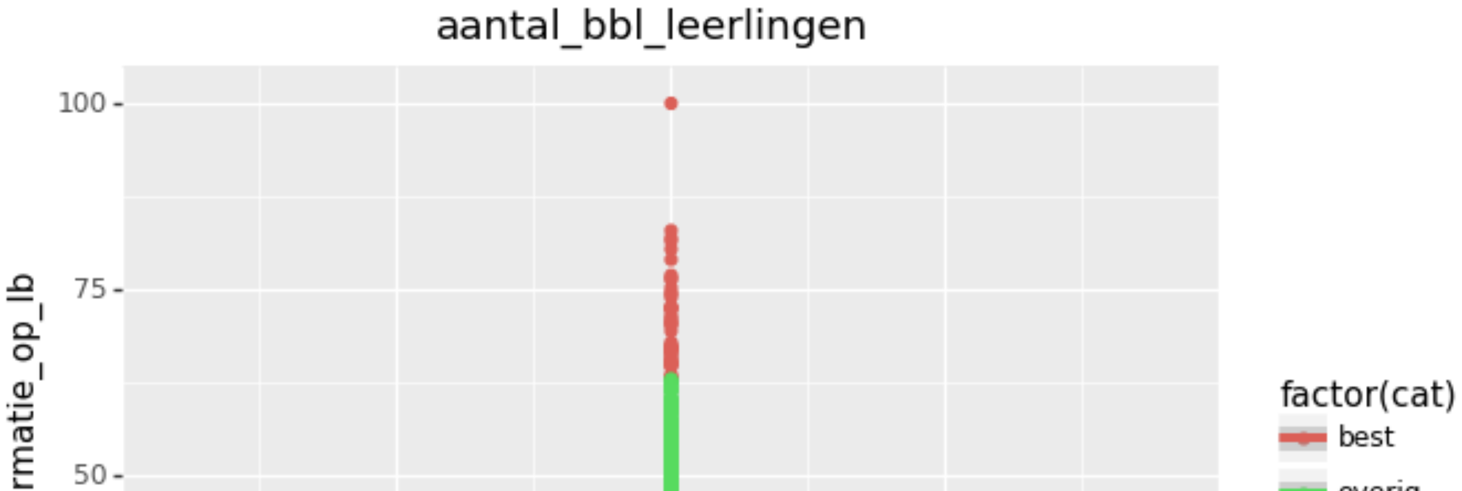


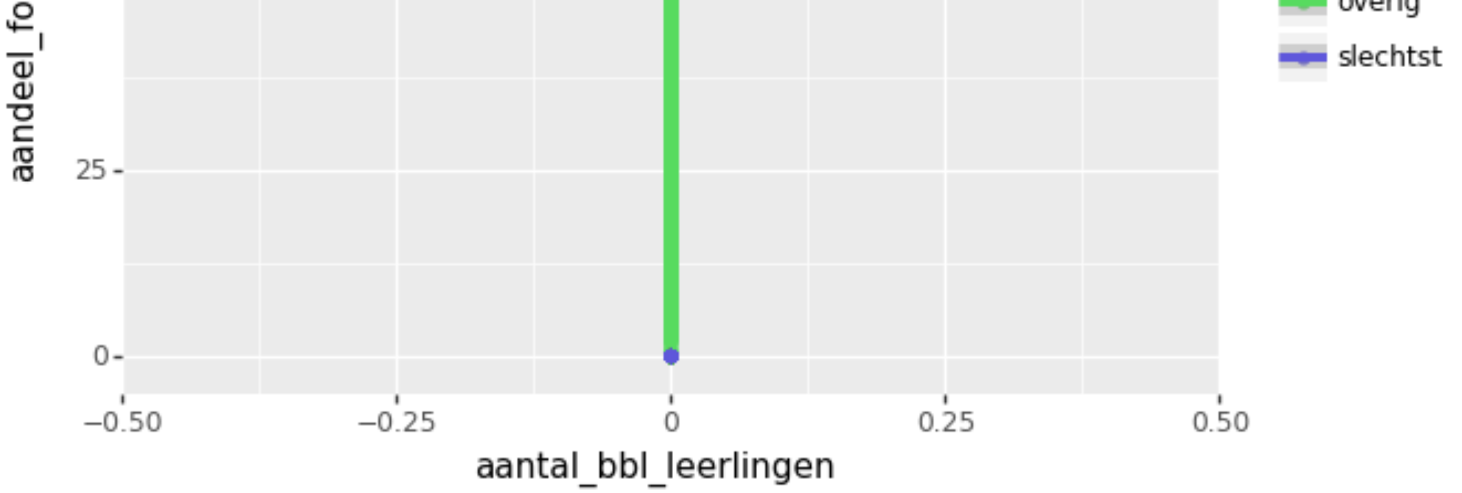


```
<ggplot: (292699312)>
aantal_bol_deeltijdleerlingen
```

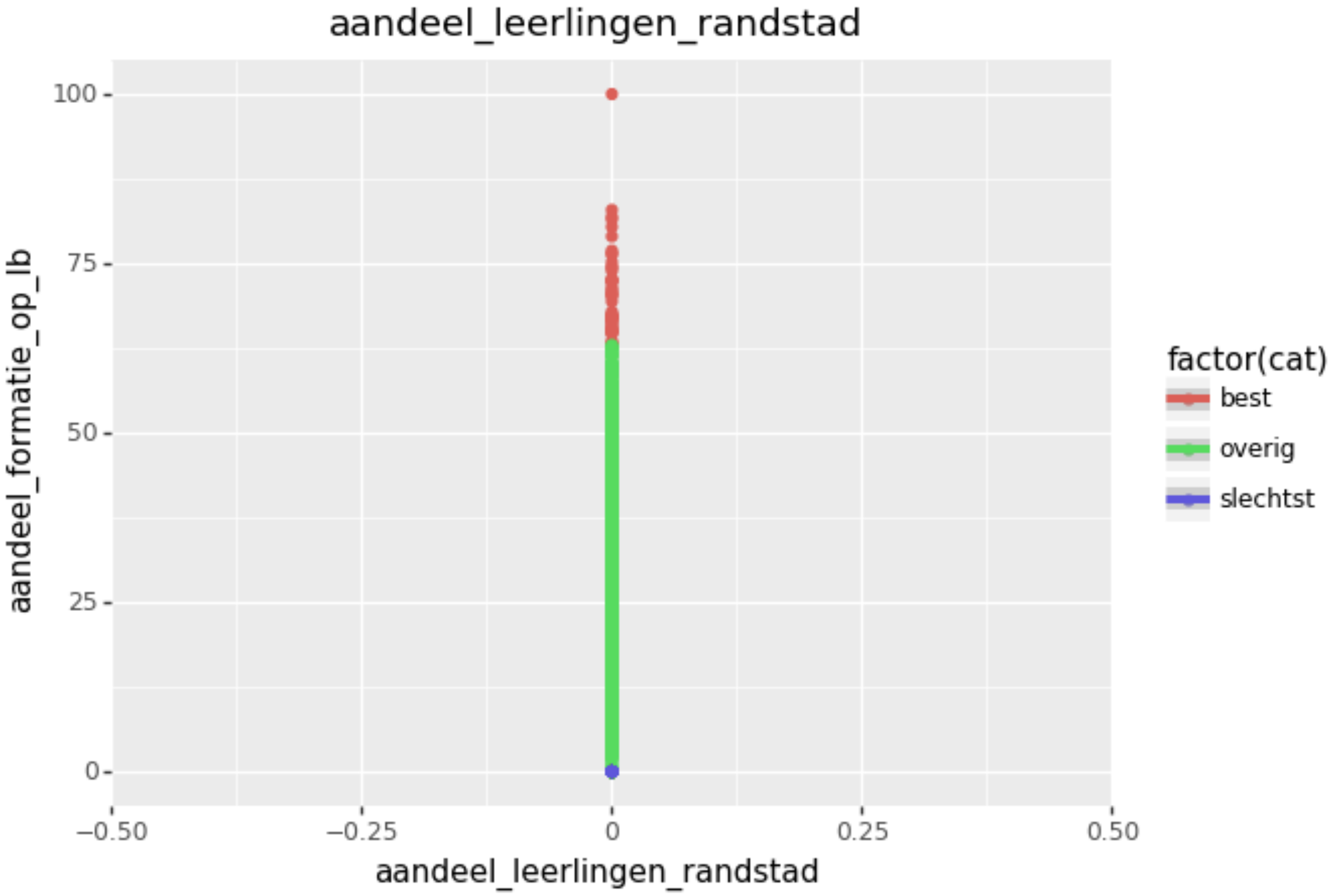


```
<ggplot: (292945135)>
aantal_bbl_leerlingen
```

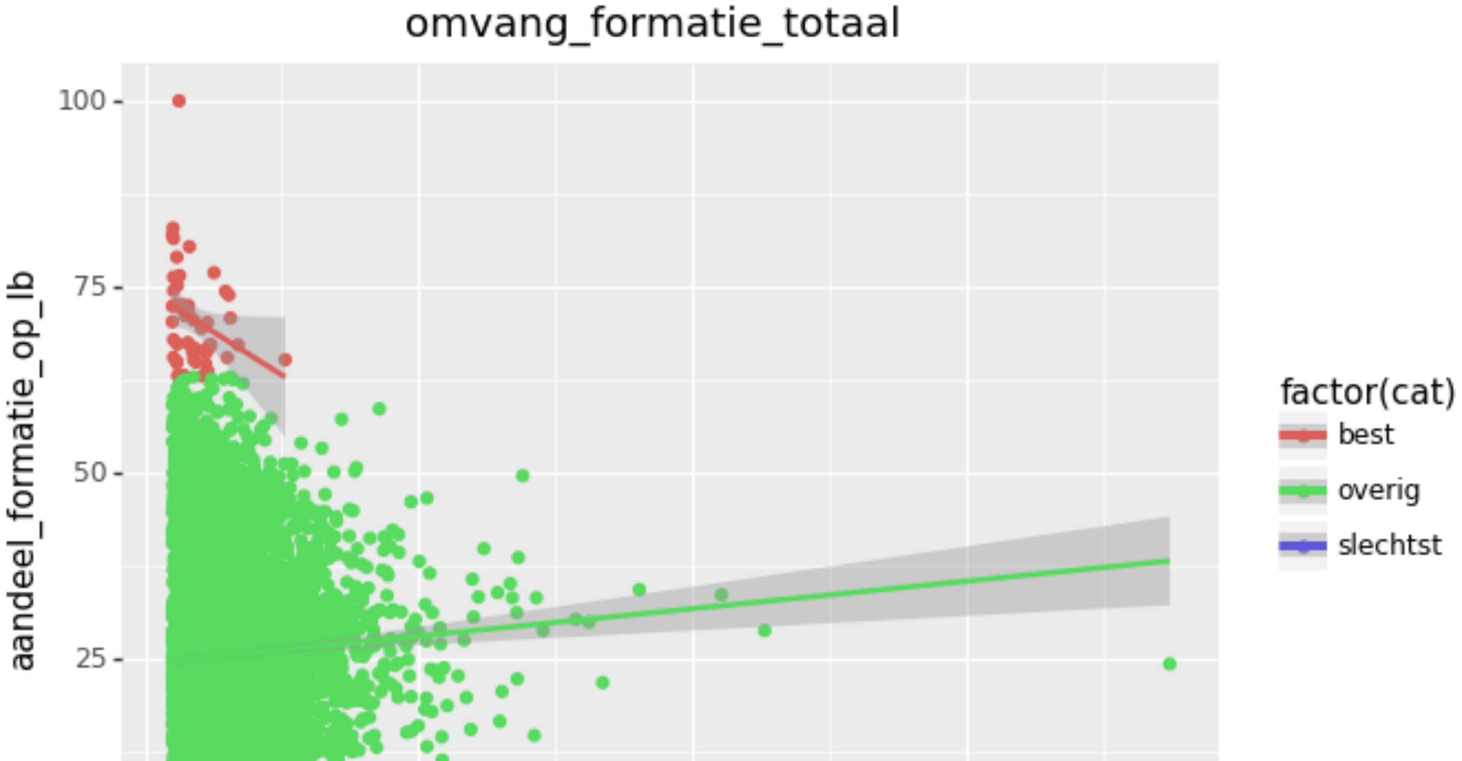


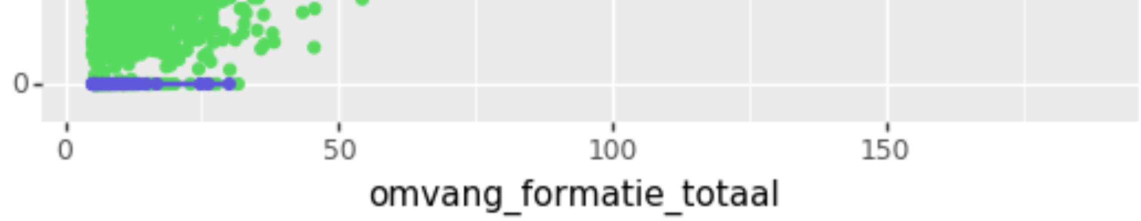


```
<ggplot: (-9223372036561454613)>
aandeel_leerlingen_randstad
```

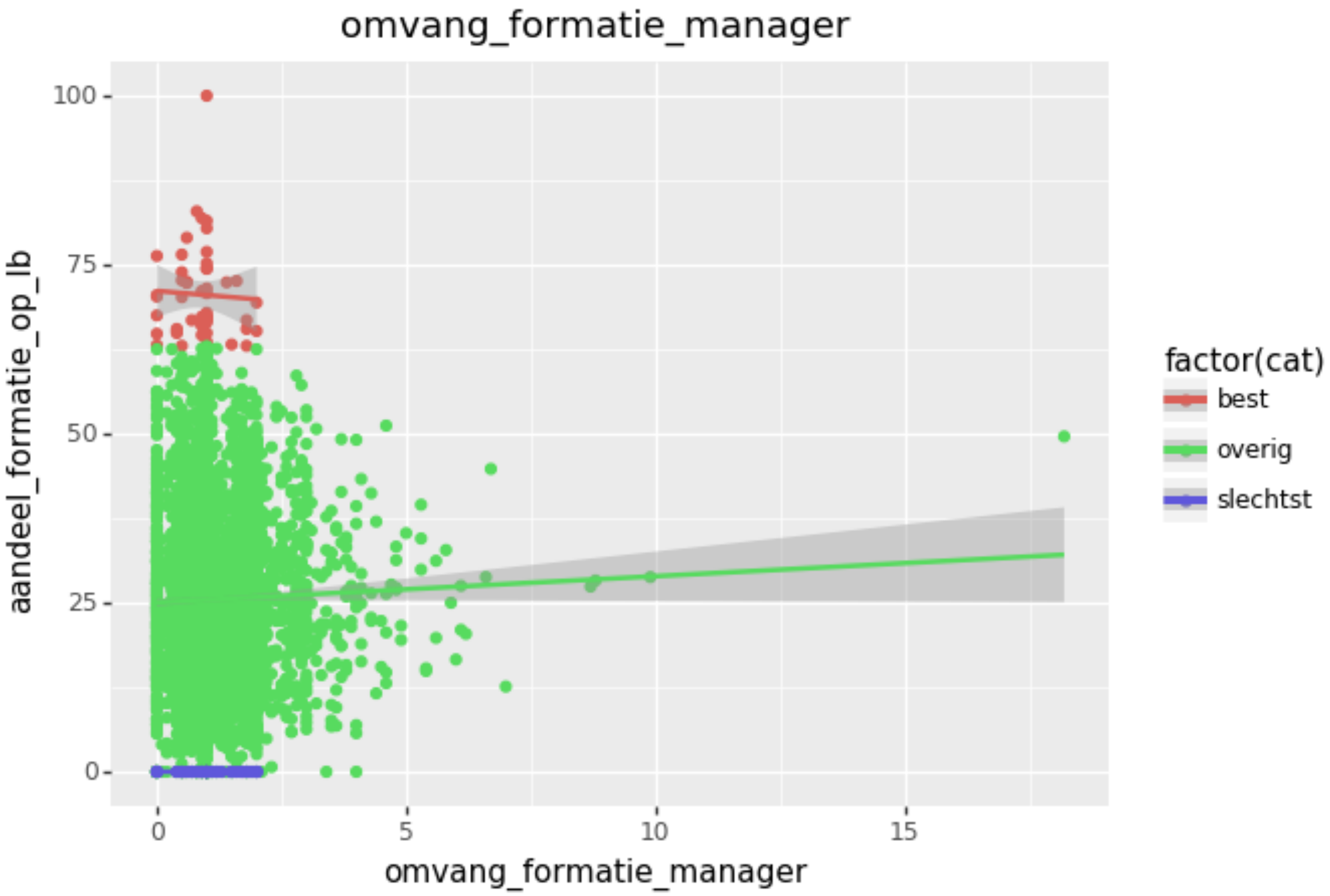


```
<ggplot: (-9223372036561175201)>
omvang_formatie_totaal
```

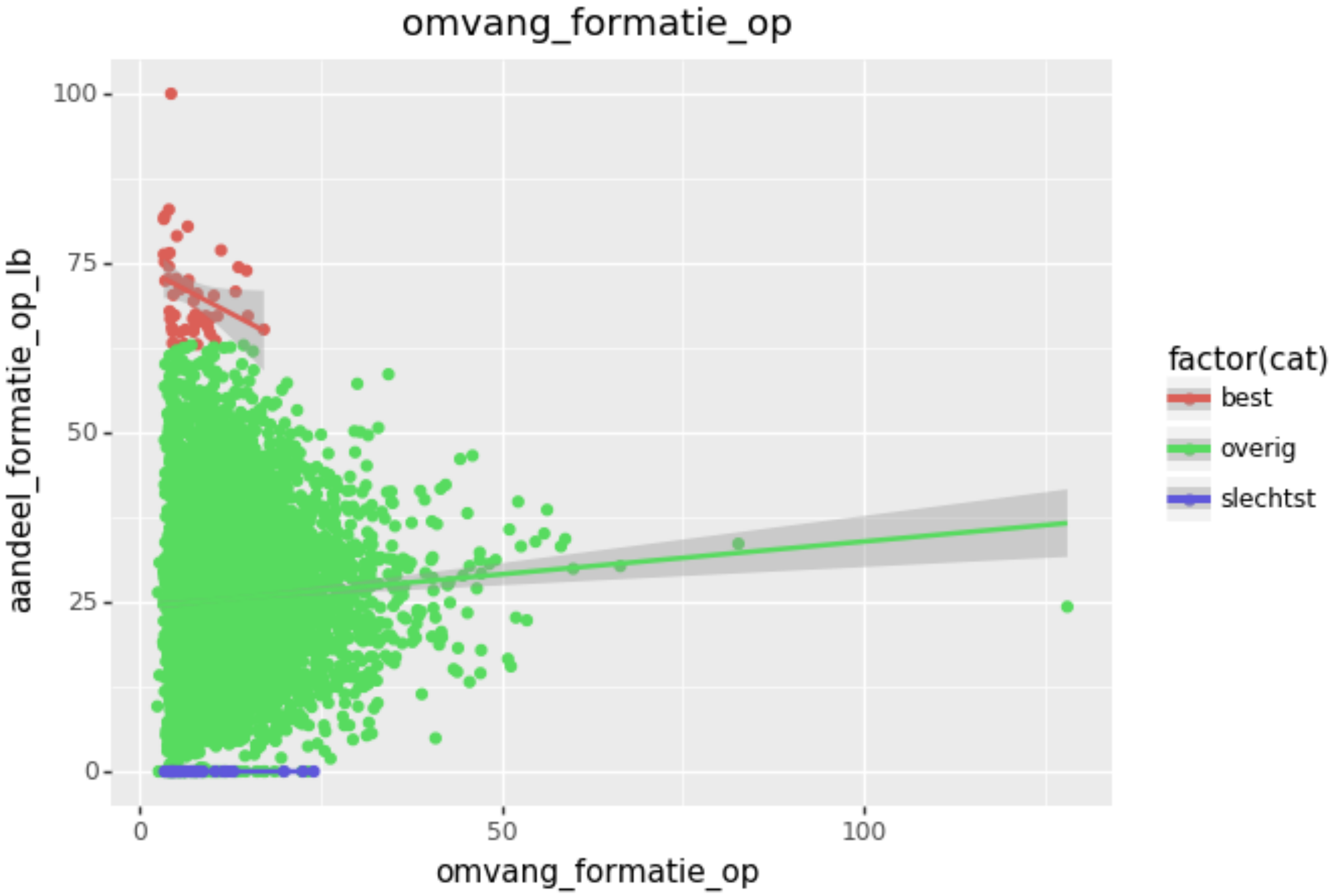




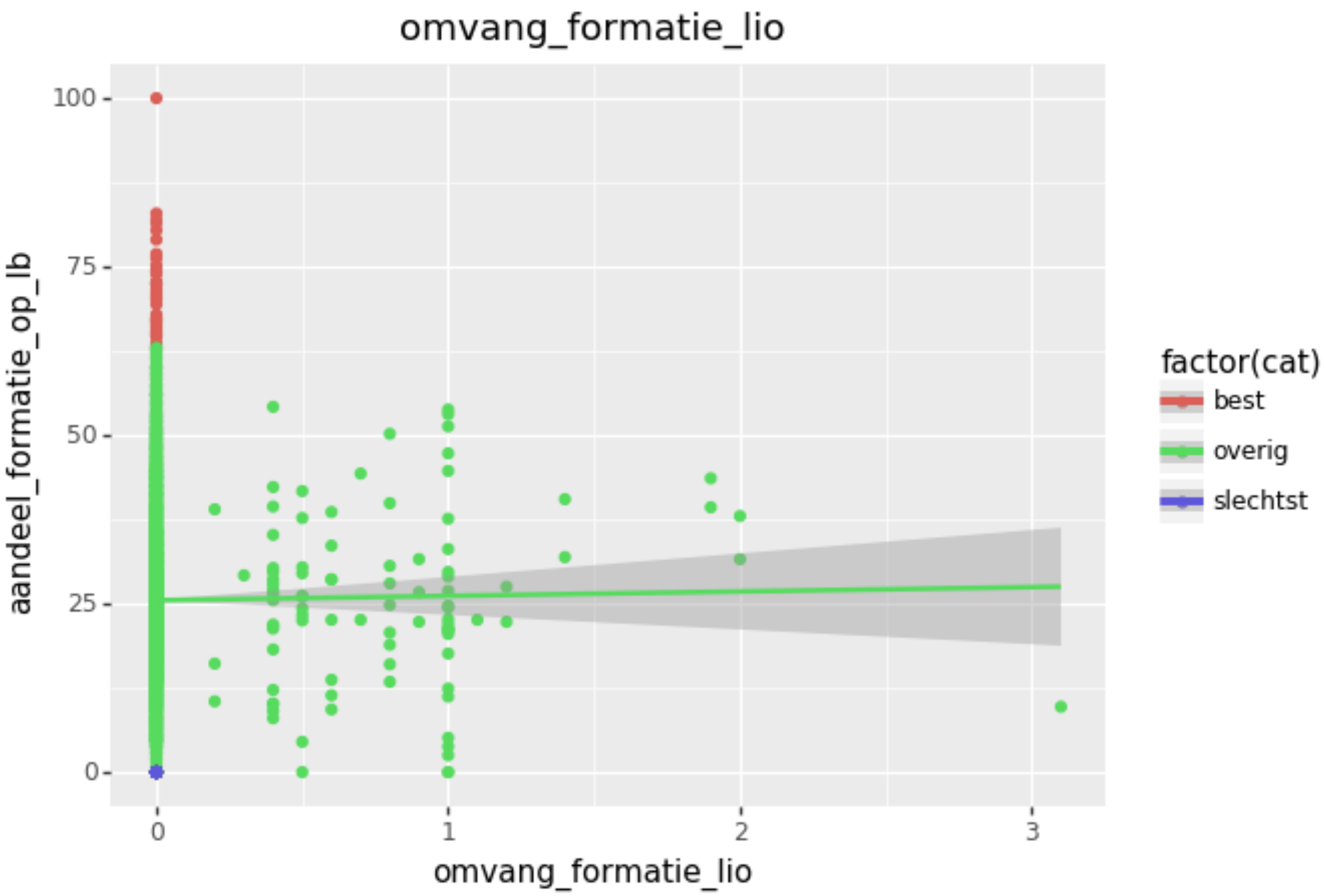
```
<ggplot: (293599289)>
omvang_formatie_manager
```



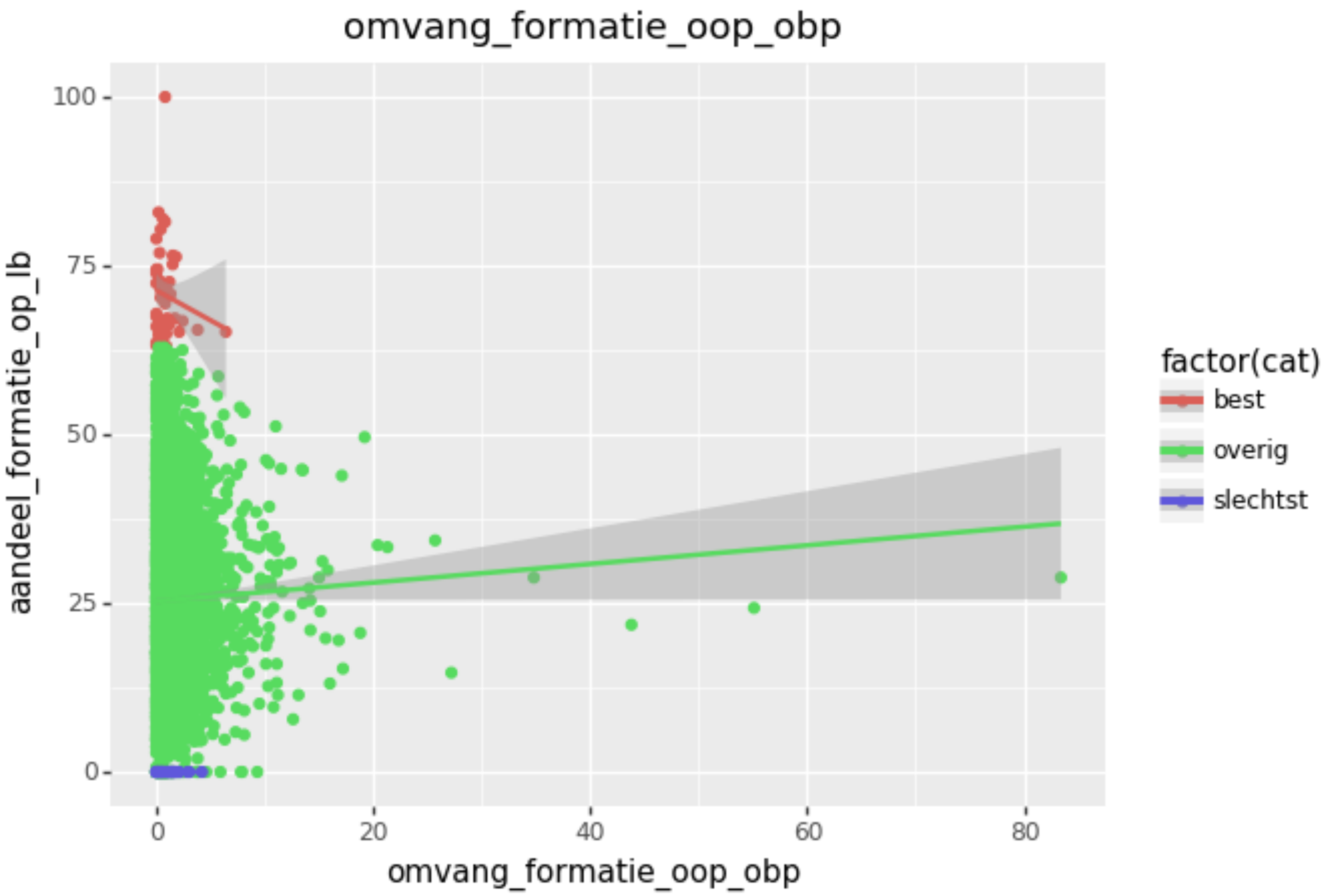
```
<ggplot: (293969100)>
omvang_formatie_op
```



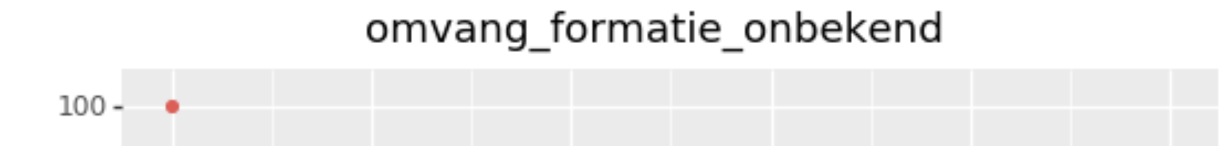
```
<ggplot: (293697021)>
omvang_formatie_lio
```

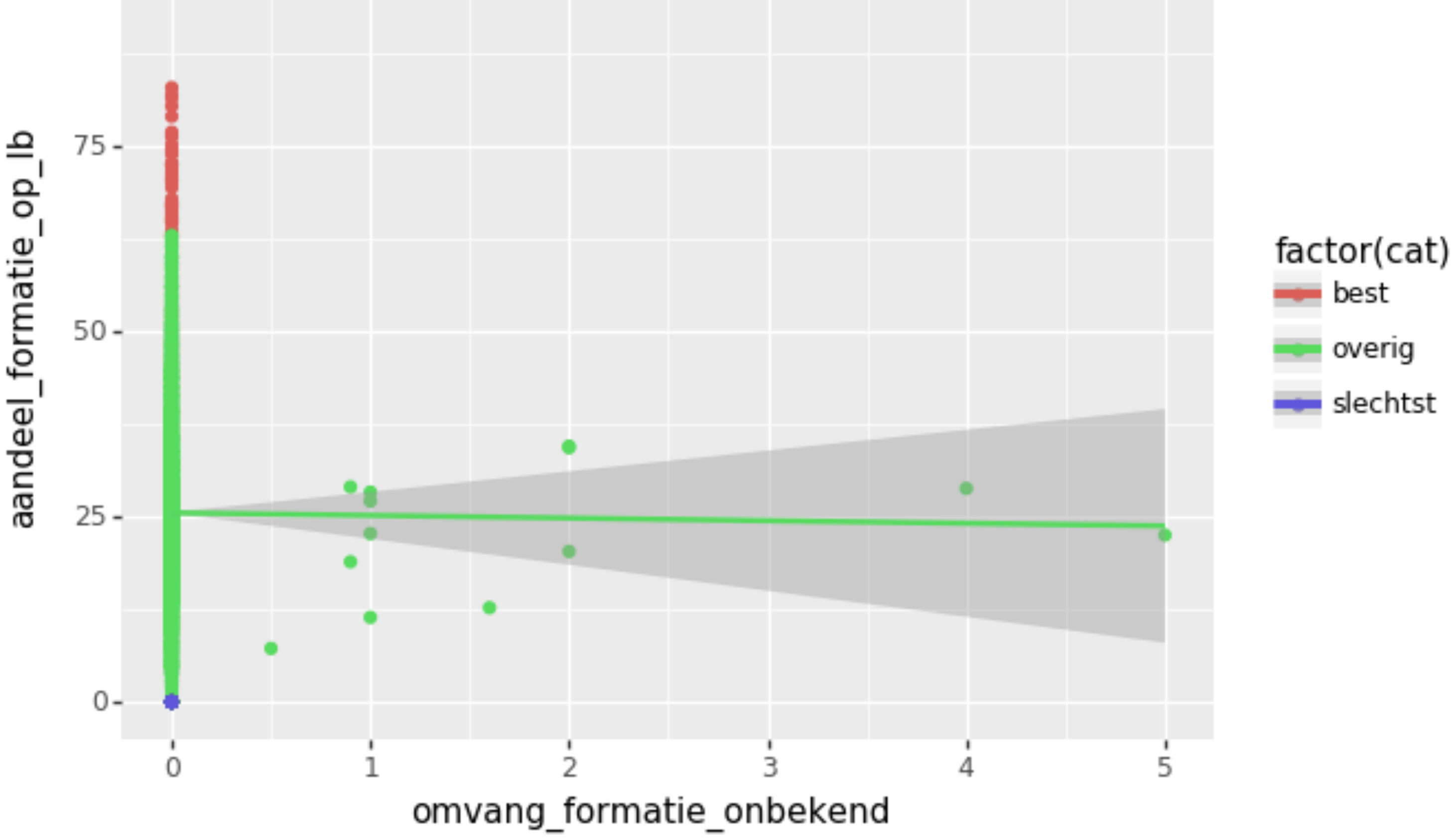


```
<ggplot: (-9223372036577143131)>
omvang_formatie_oop_obp
```

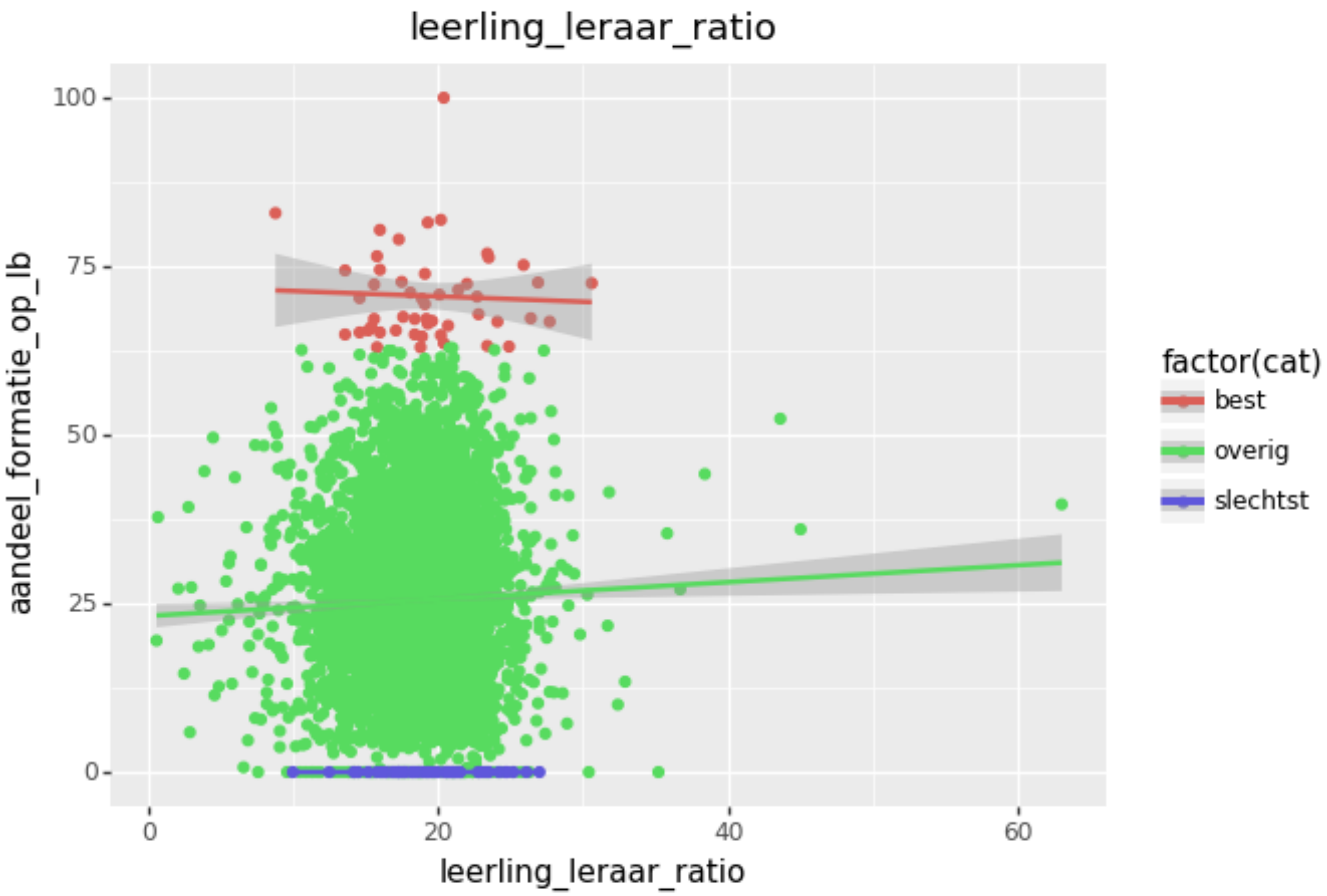


```
<ggplot: (285906026)>
omvang_formatie_onbekend
```

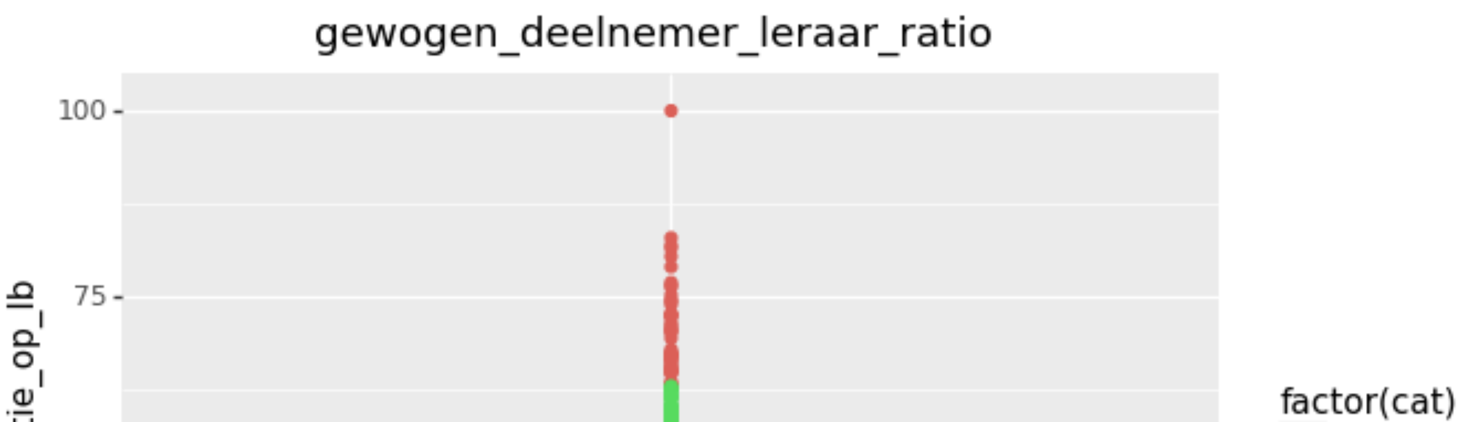


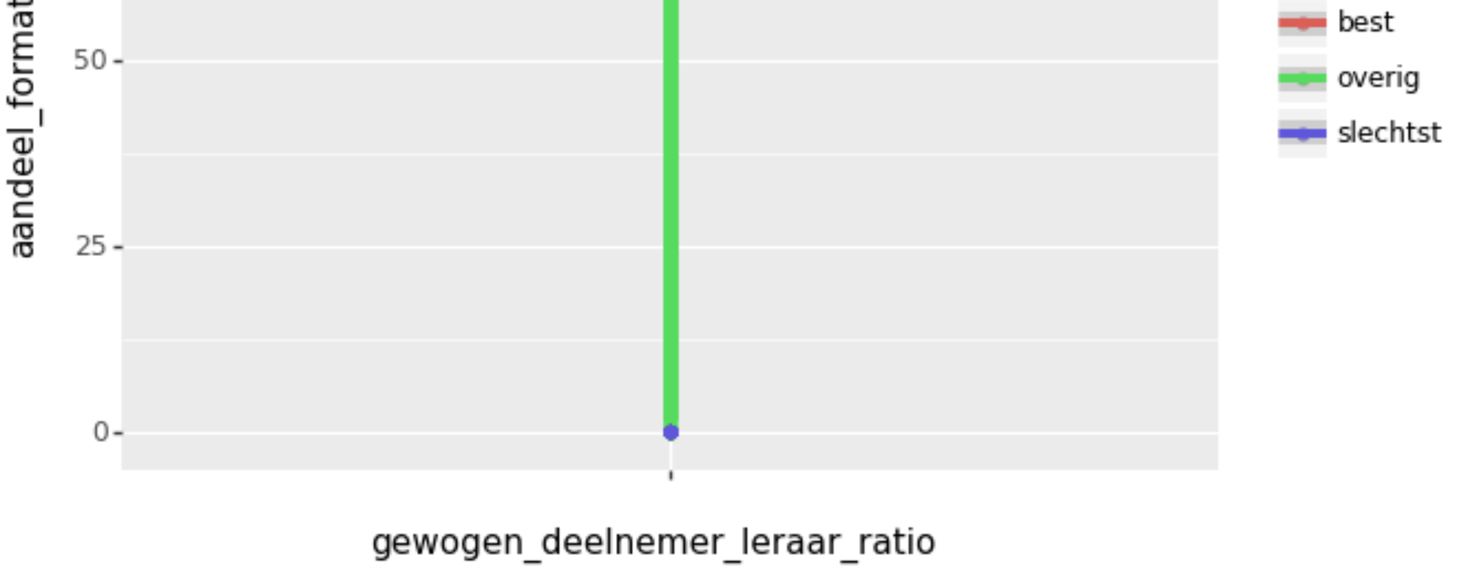


```
<ggplot: (-9223372036568861359)>
leerling_leraar_ratio
```

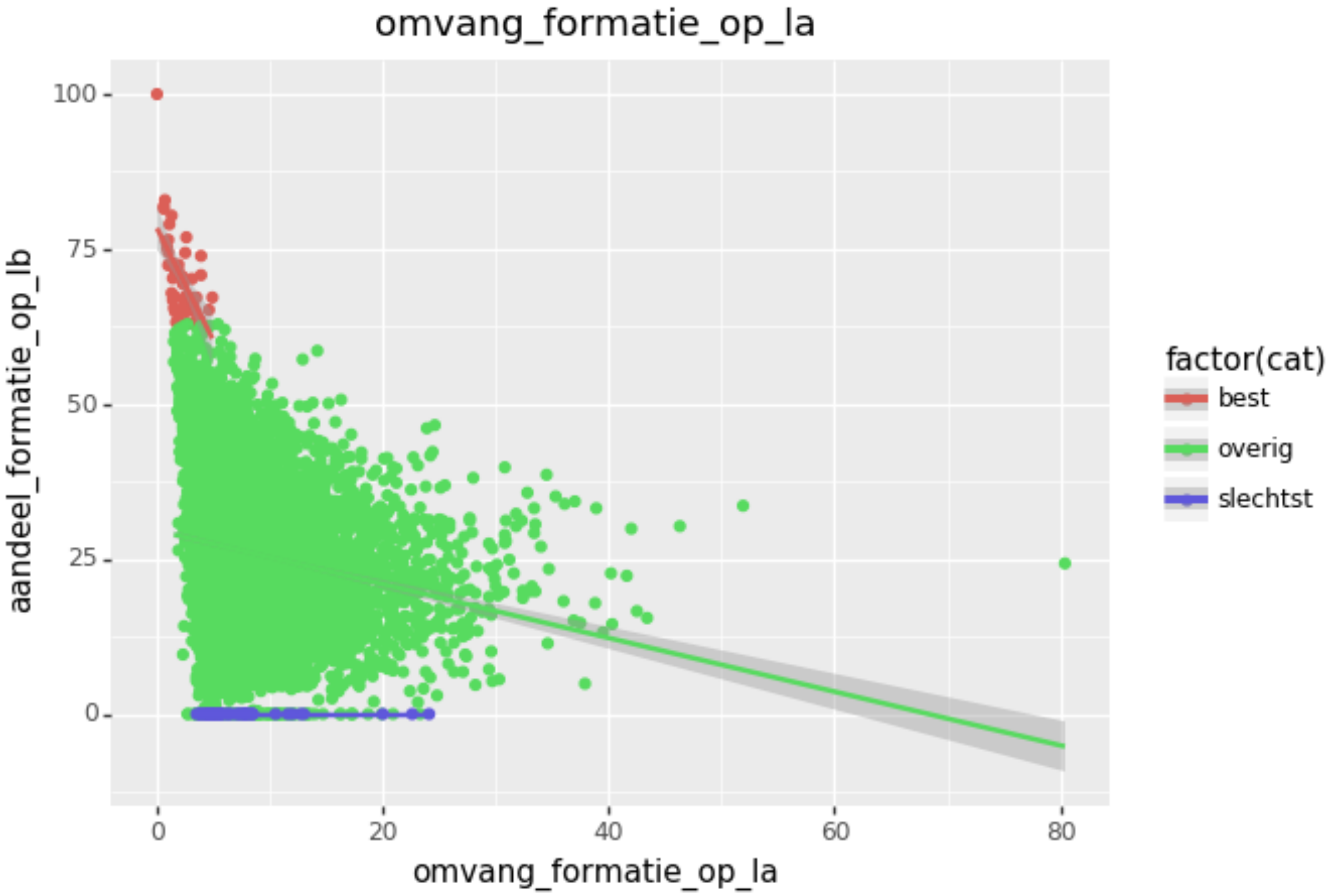


```
<ggplot: (286736185)>
gewogen_deelnemer_leraar_ratio
```

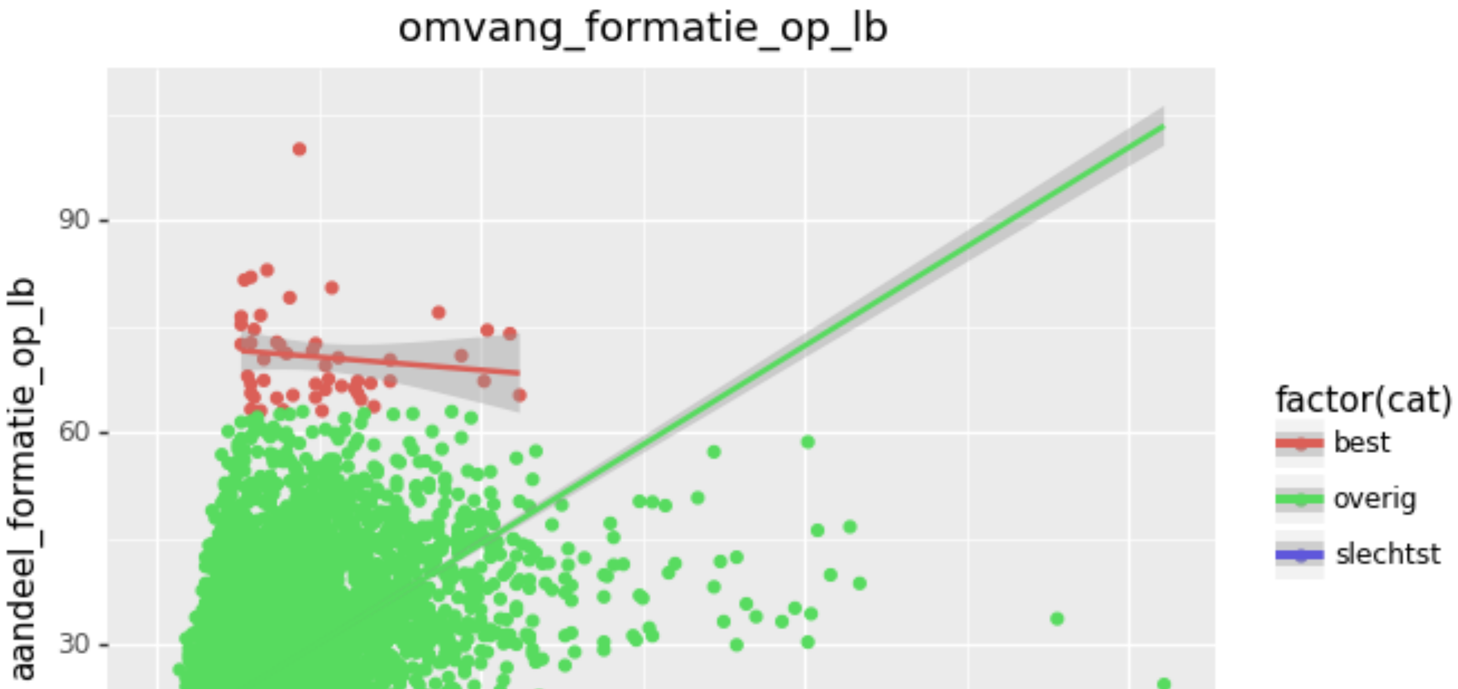


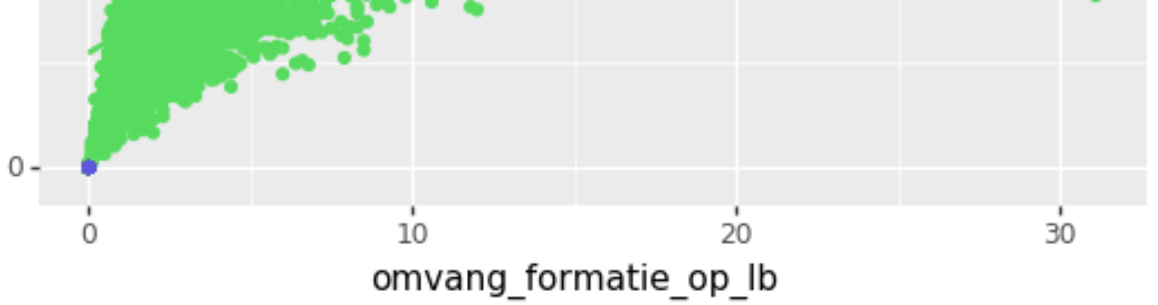


```
<ggplot: (-9223372036566279687)>
omvang_formatie_op_la
```

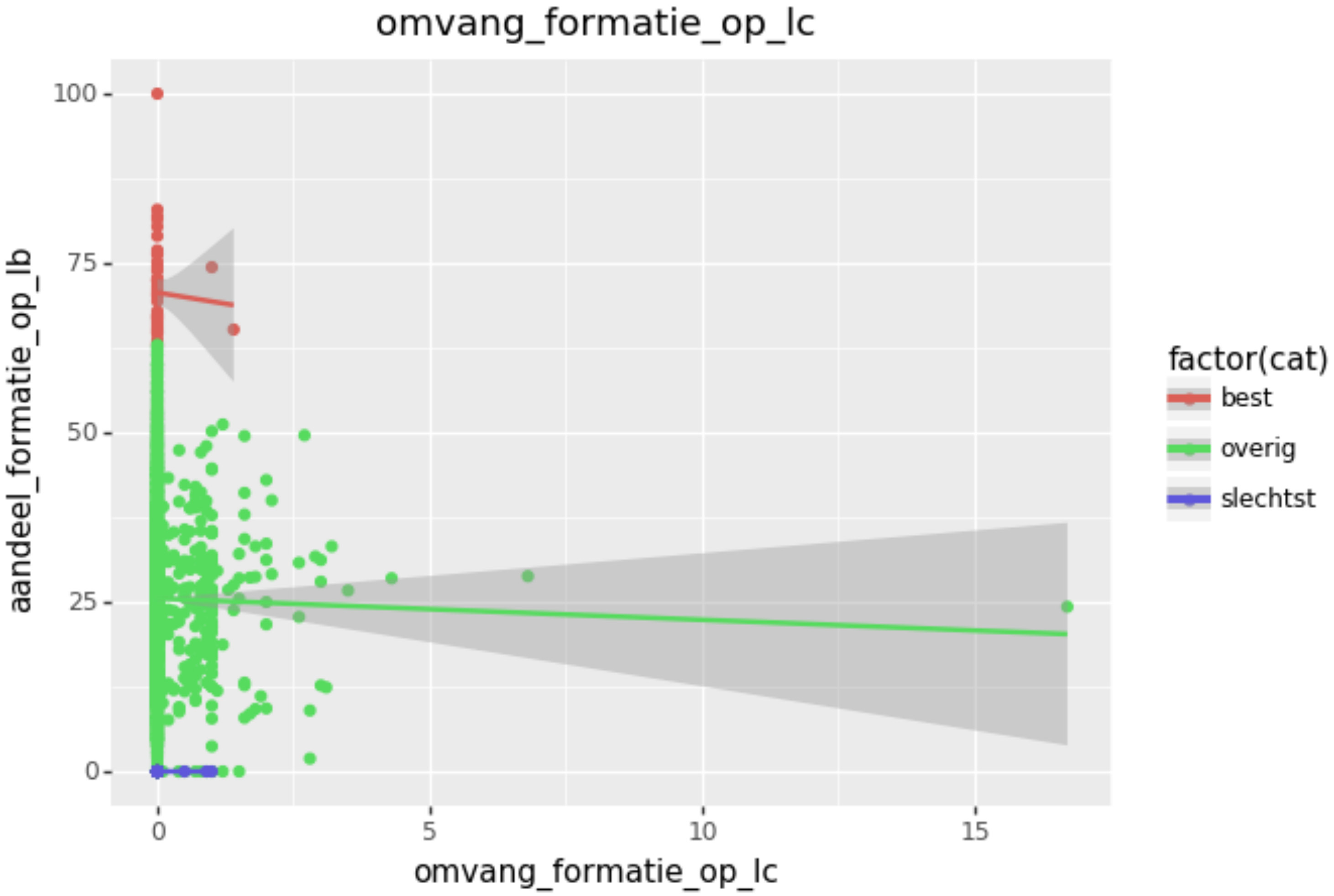


```
<ggplot: (-9223372036565482878)>
omvang_formatie_op_lb
```

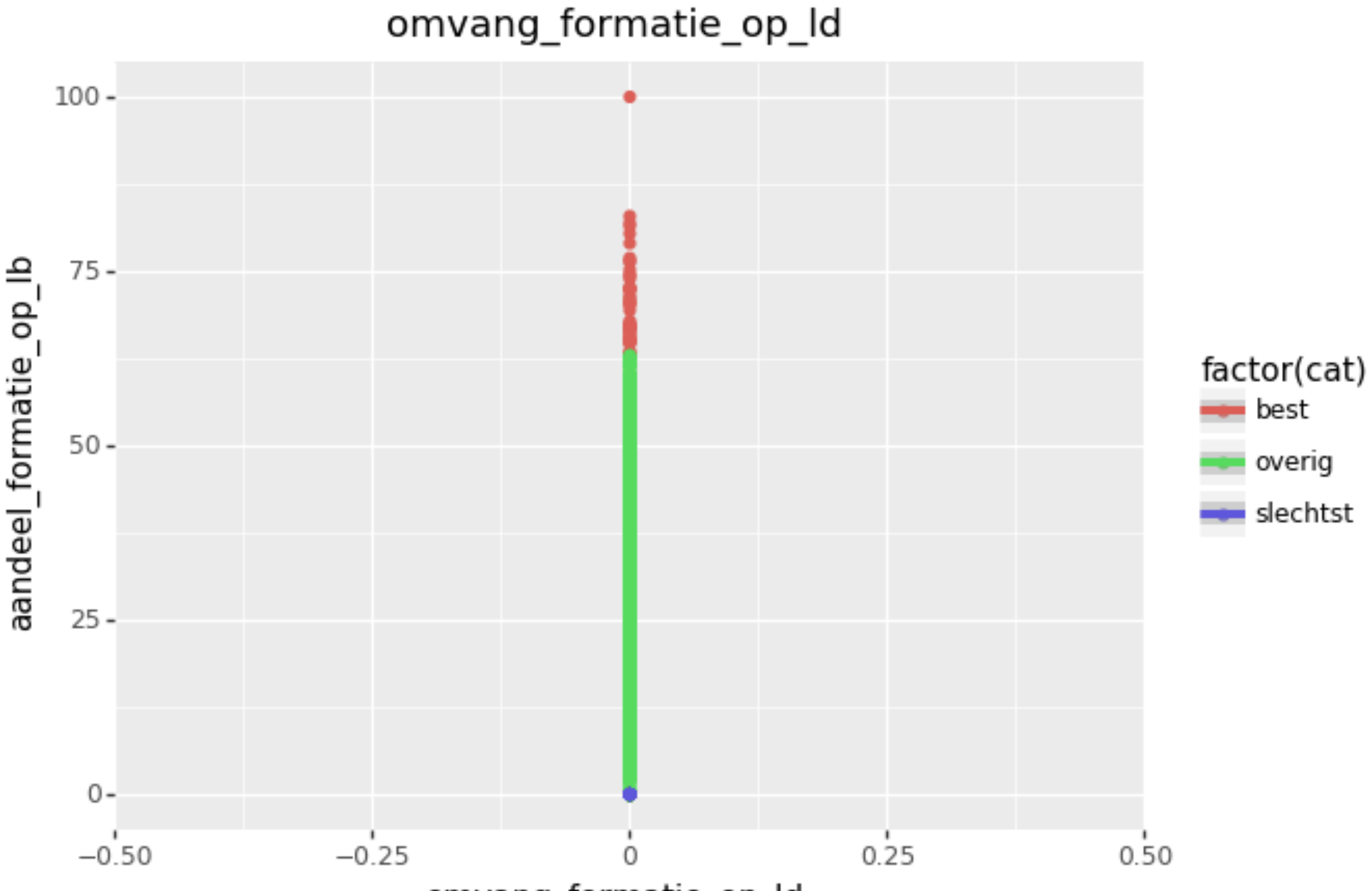


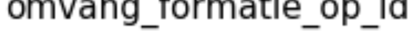


```
<ggplot: (-9223372036567815189)>
omvang_formatie_op_lc
```

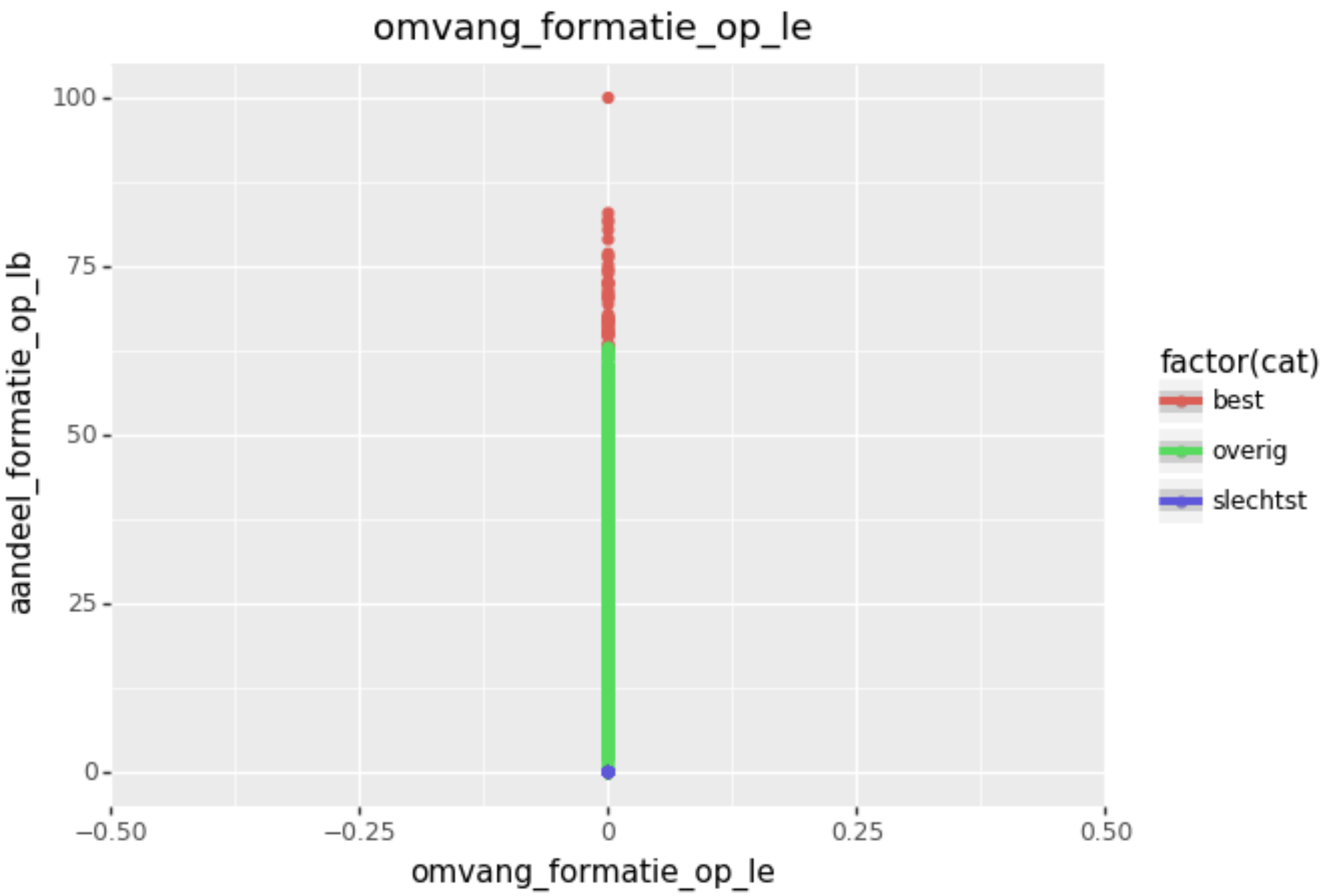


```
<ggplot: (292179832)>
omvang_formatie_op_ld
```

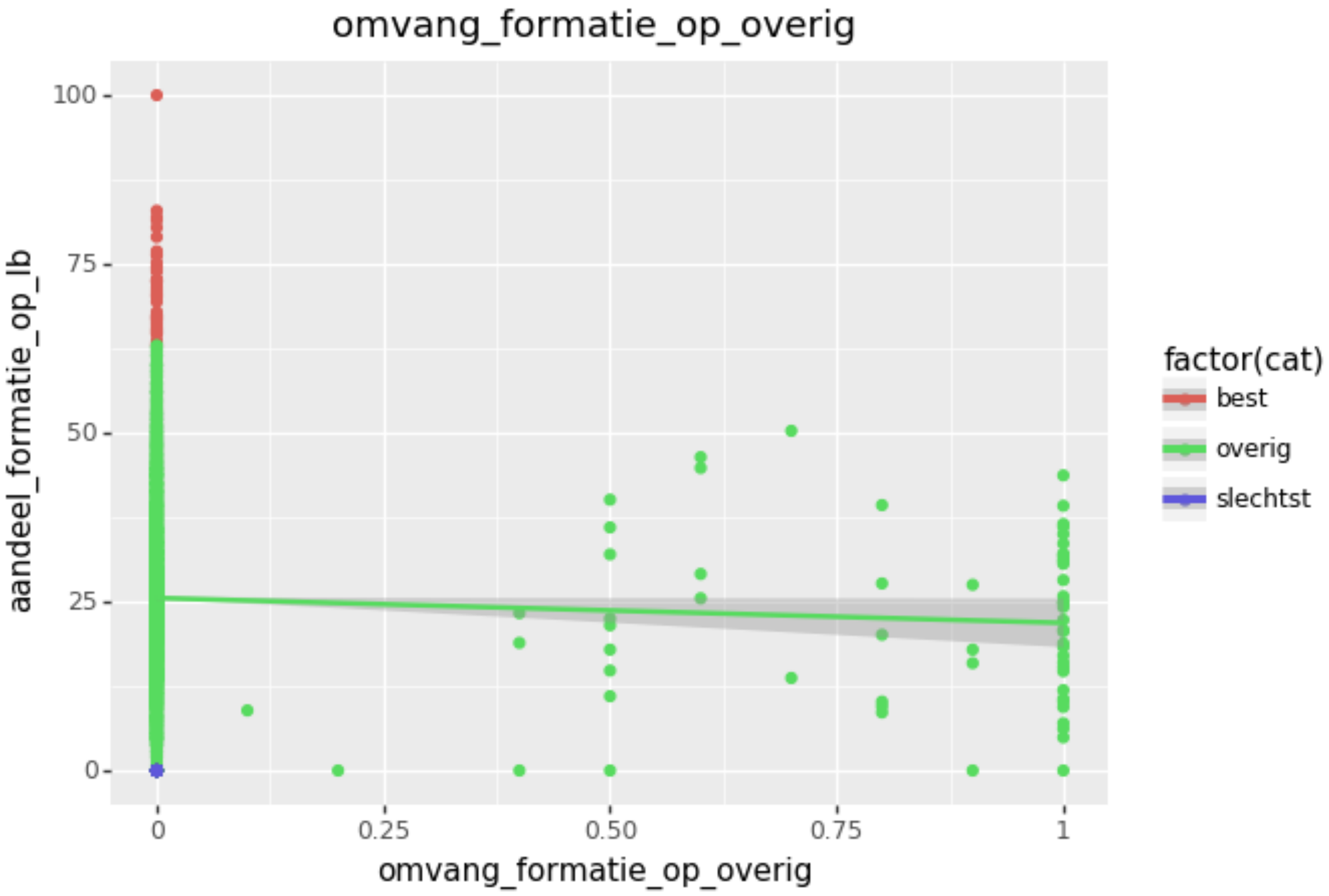




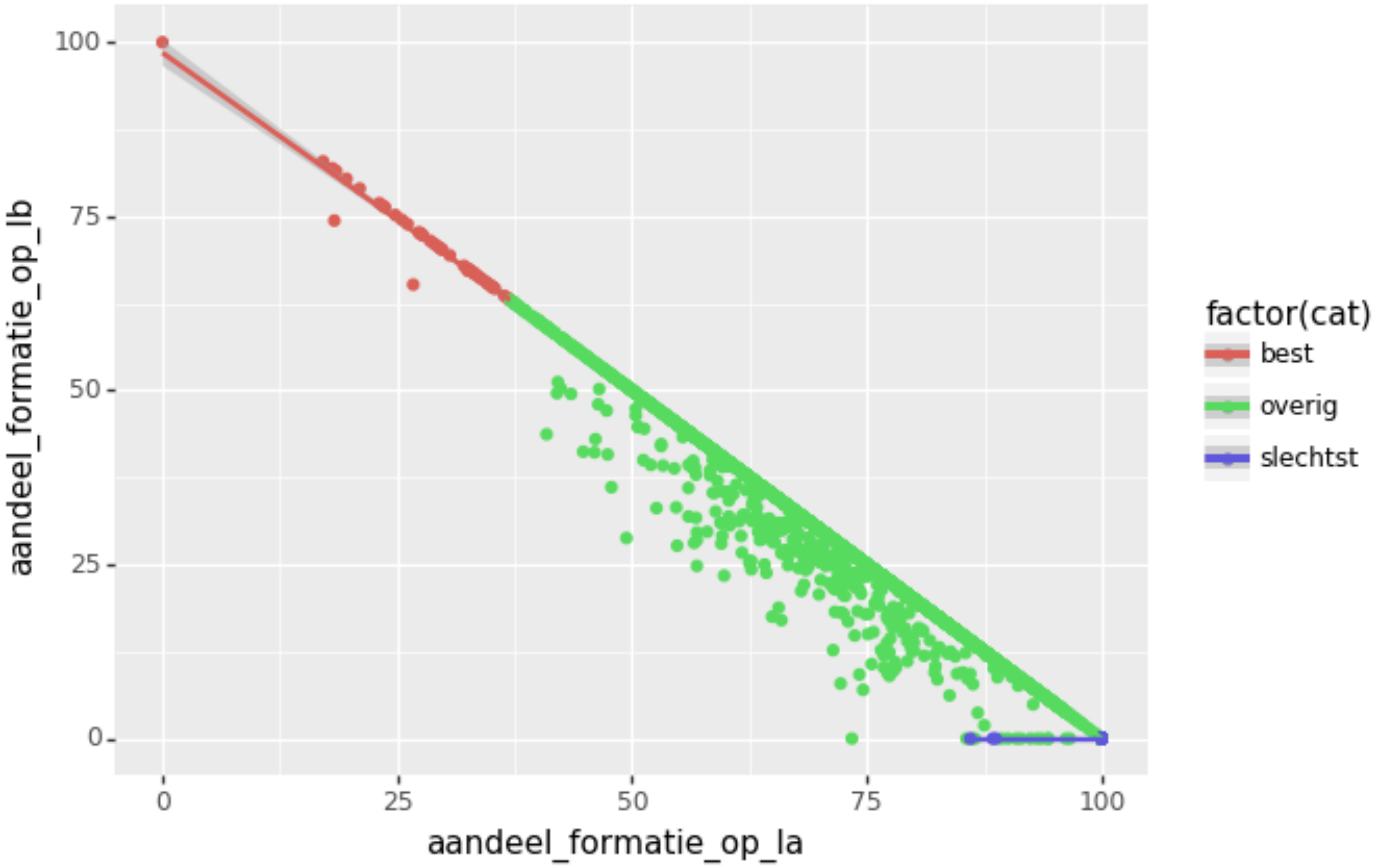
```
<ggplot: (286732594)>
omvang_formatie_op_le
```



```
<ggplot: (-9223372036564719847)>
omvang_formatie_op_overig
```

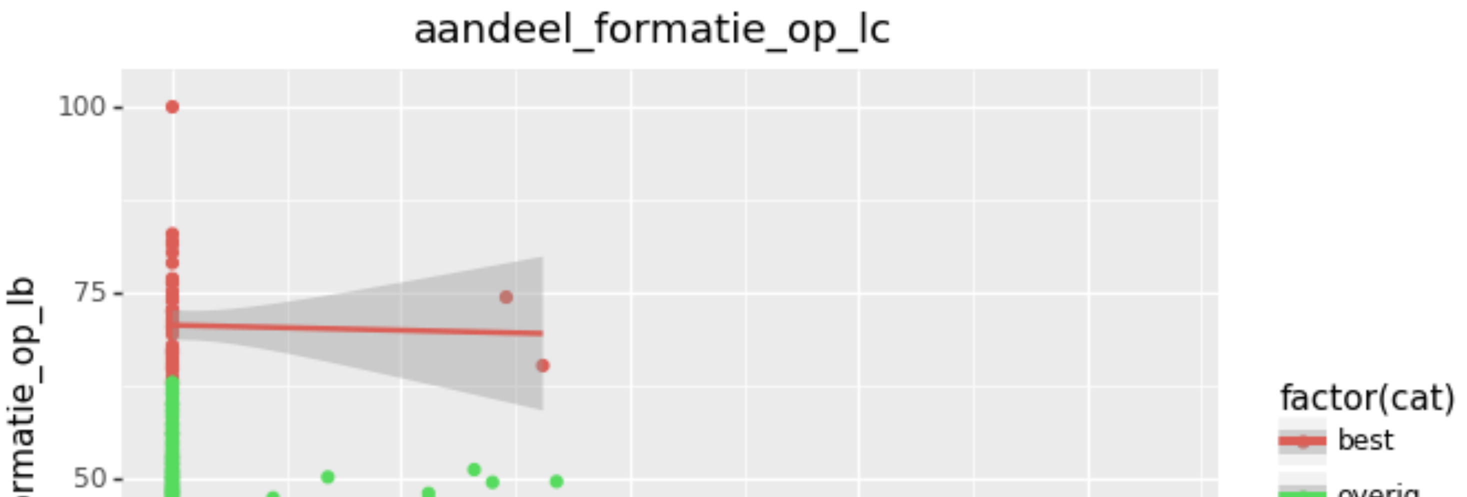


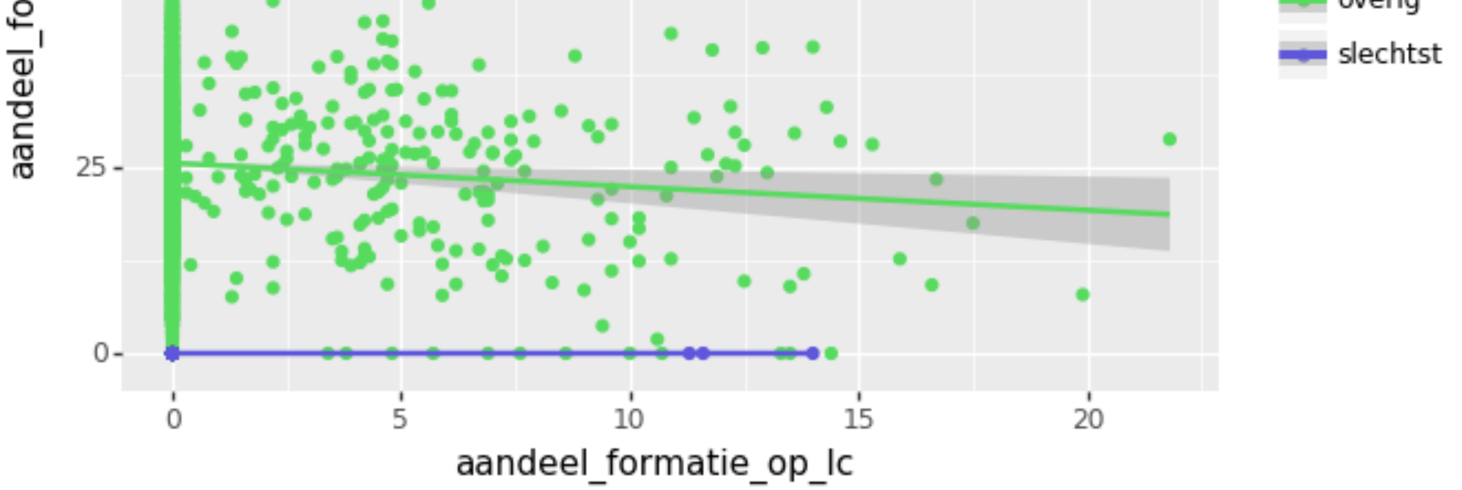
```
<ggplot: (293591763)>
aandeel_formatie_op_la
```



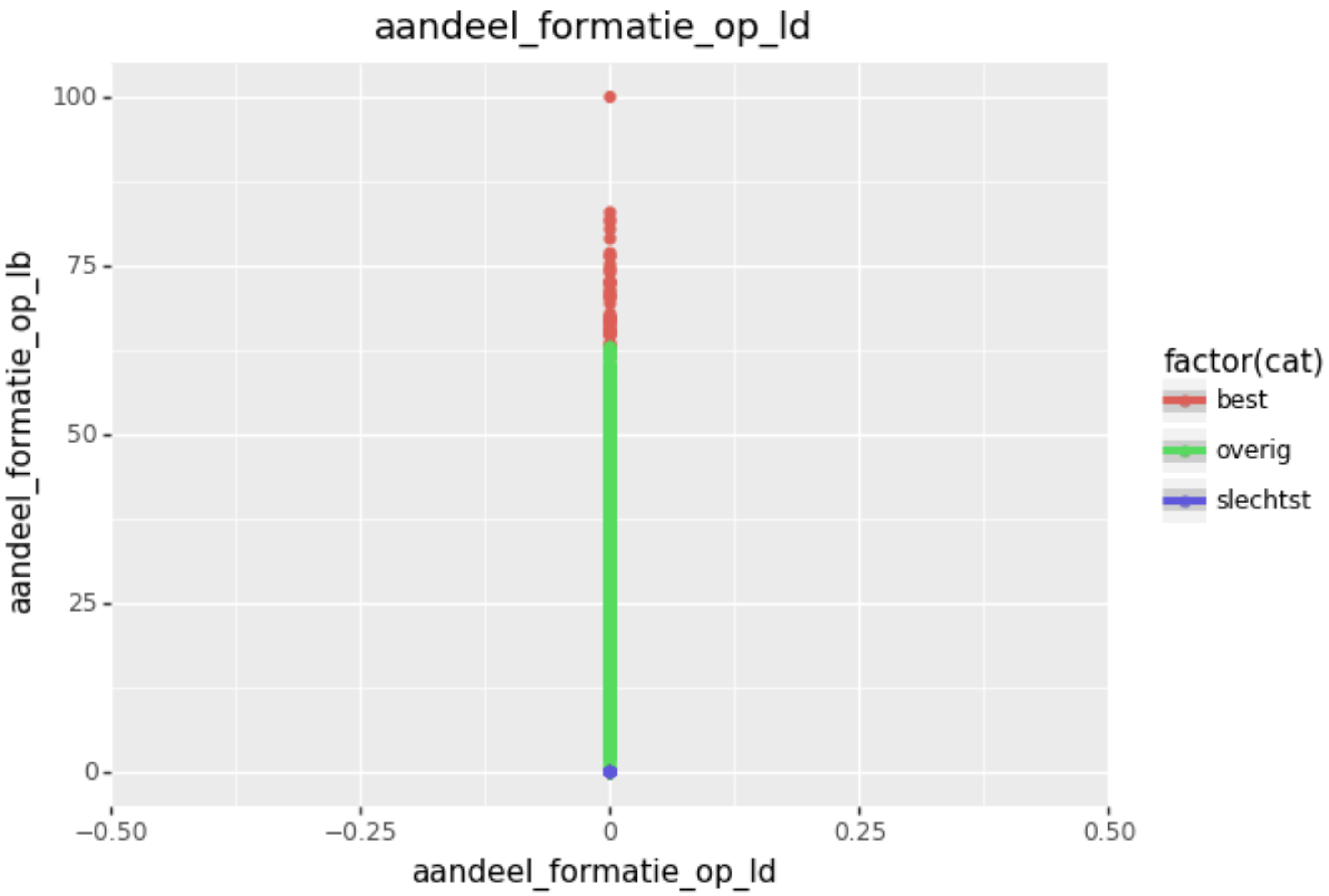
```
<ggplot: (-9223372036568014178)>
aandeel_formatie_op_lb
Wrong number of items passed 2, placement implies 1
aandeel_formatie_op_lc

Traceback (most recent call last):
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/indexes/base.py", line 2525, in get_loc
    return self._engine.get_loc(key)
  File "pandas/_libs/index.pyx", line 117, in pandas._libs.index.Ind
exEngine.get_loc
  File "pandas/_libs/index.pyx", line 139, in pandas._libs.index.Ind
exEngine.get_loc
  File "pandas/_libs/hashtable_class_helper.pxi", line 1265, in pand
as._libs.hashtable.PyObjectHashTable.get_item
  File "pandas/_libs/hashtable_class_helper.pxi", line 1273, in pand
as._libs.hashtable.PyObjectHashTable.get_item
KeyError: 'x'

During handling of the above exception, another exception occurred:

Traceback (most recent call last):
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/internals.py", line 3968, in set
    loc = self.items.get_loc(item)
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/indexes/base.py", line 2527, in get_loc
    return self._engine.get_loc(self._maybe_cast_indexer(key))
  File "pandas/_libs/index.pyx", line 117, in pandas._libs.index.Ind
exEngine.get_loc
  File "pandas/_libs/index.pyx", line 139, in pandas._libs.index.Ind
exEngine.get_loc
  File "pandas/_libs/hashtable_class_helper.pxi", line 1265, in pand
as._libs.hashtable.PyObjectHashTable.get_item
  File "pandas/_libs/hashtable_class_helper.pxi", line 1273, in pand
as._libs.hashtable.PyObjectHashTable.get_item
```

```
KeyError: 'x'

During handling of the above exception, another exception occurred:

Traceback (most recent call last):
  File "<ipython-input-1-0789d905e37e>", line 69, in <module>
    + stat_smooth(method='lm'))
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/plotnine/ggplot.py", line 86, in __repr__
    self.draw()
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/plotnine/ggplot.py", line 179, in draw
    self._build()
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/plotnine/ggplot.py", line 276, in _build
    layers.compute_aesthetics(self)
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/plotnine/layer.py", line 82, in compute_aesthetics
    l.compute_aesthetics(plot)
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/plotnine/layer.py", line 308, in compute_aesthetics
    evaled[ae] = data[col]
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/frame.py", line 2519, in __setitem__
    self._set_item(key, value)
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/frame.py", line 2586, in _set_item
    NDFrame._set_item(self, key, value)
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/generic.py", line 1954, in _set_item
    self._data.set(key, value)
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/internals.py", line 3971, in set
    self.insert(len(self.items), item, value)
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/internals.py", line 4072, in insert
    placement=slice(loc, loc + 1))
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/internals.py", line 2957, in make_block
    return klass(values, ndim=ndim, fastpath=fastpath, placement=pla
cement)
  File "/Users/rix0rrr/Dev/CorrespondentPO/virtualenv/lib/python3.6/
site-packages/pandas/core/internals.py", line 120, in __init__
    len(self.mgr_locs)))
ValueError: Wrong number of items passed 2, placement implies 1
```

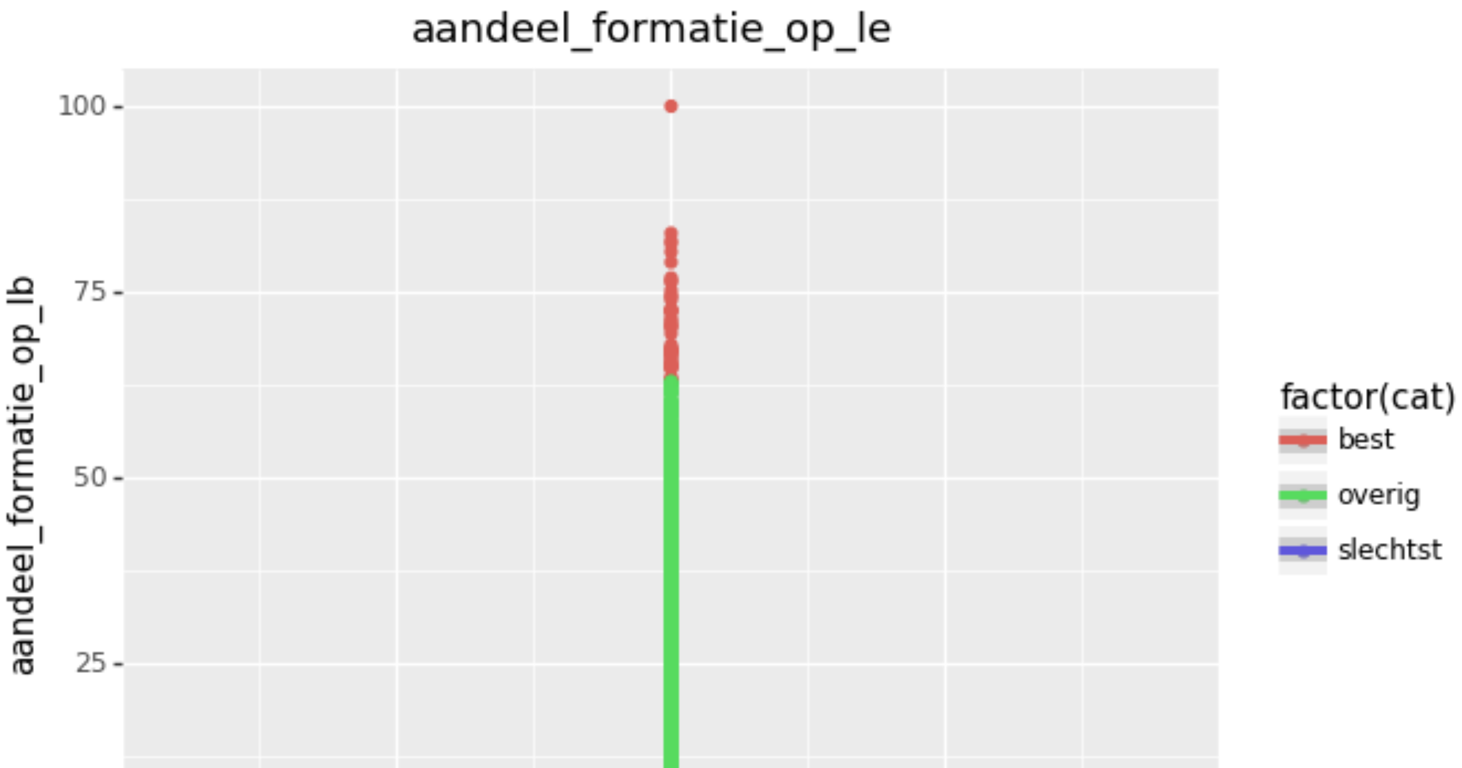


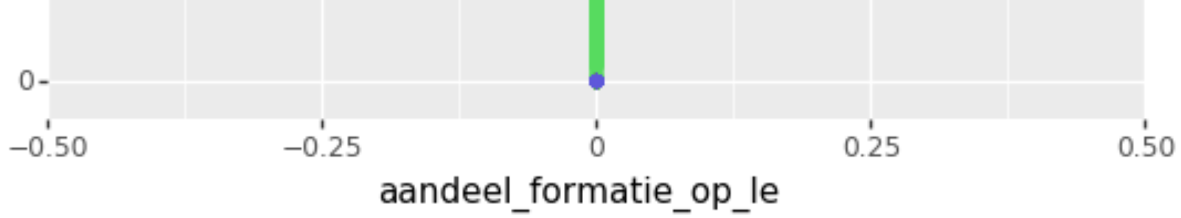


```
<ggplot: (291391990)>
aandeel_formatie_op_ld
```

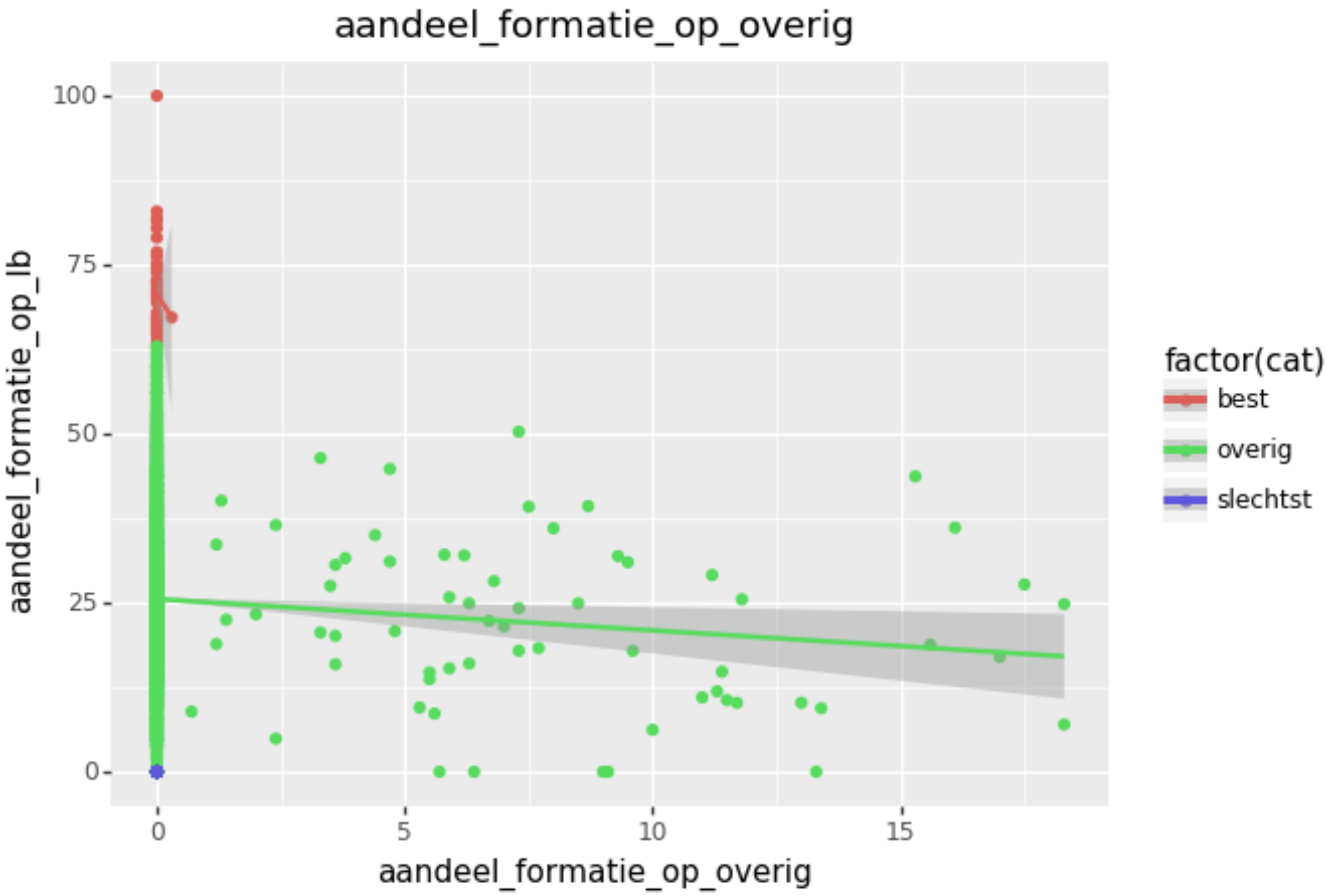


```
<ggplot: (-9223372036563381907)>
aandeel_formatie_op_le
```

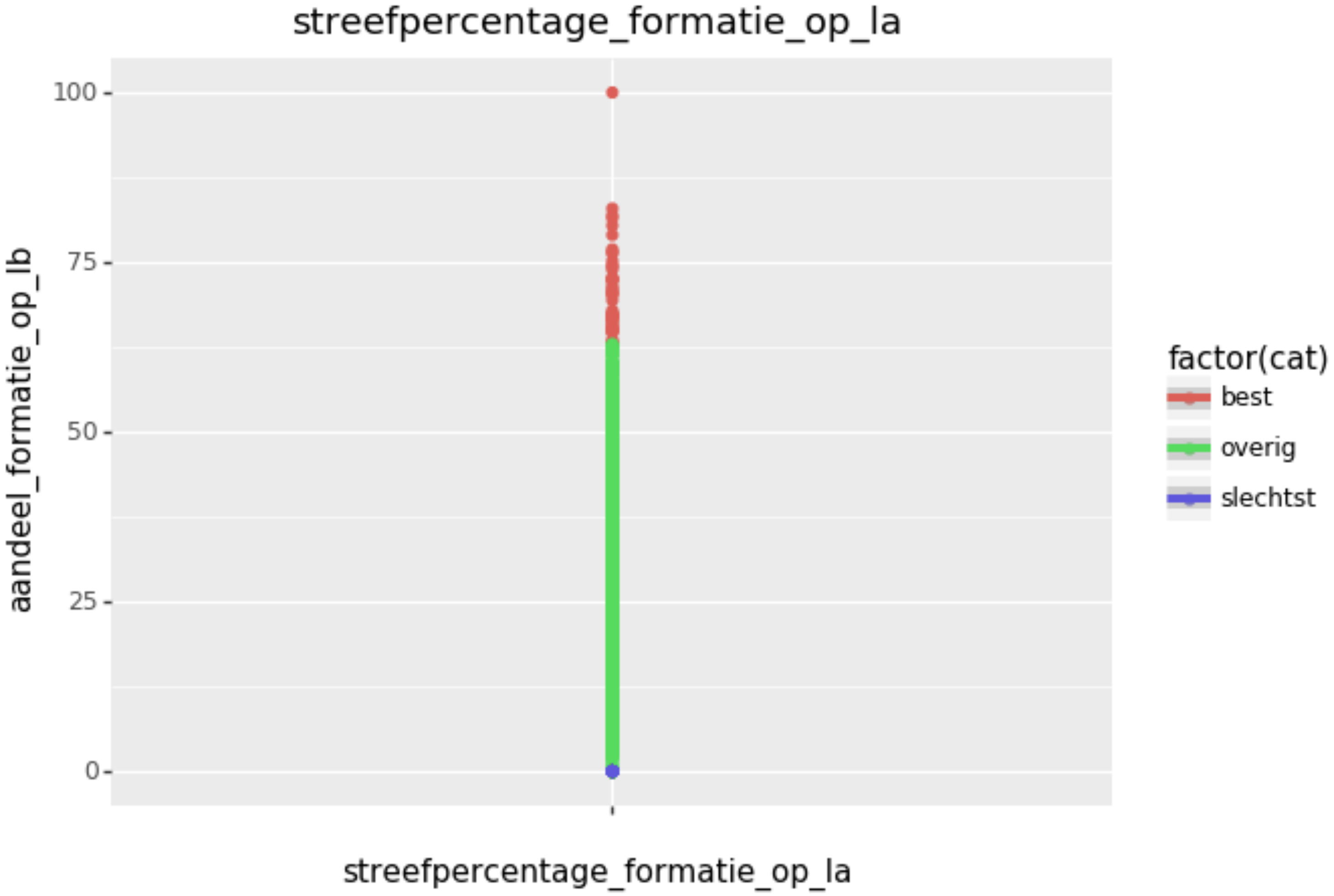




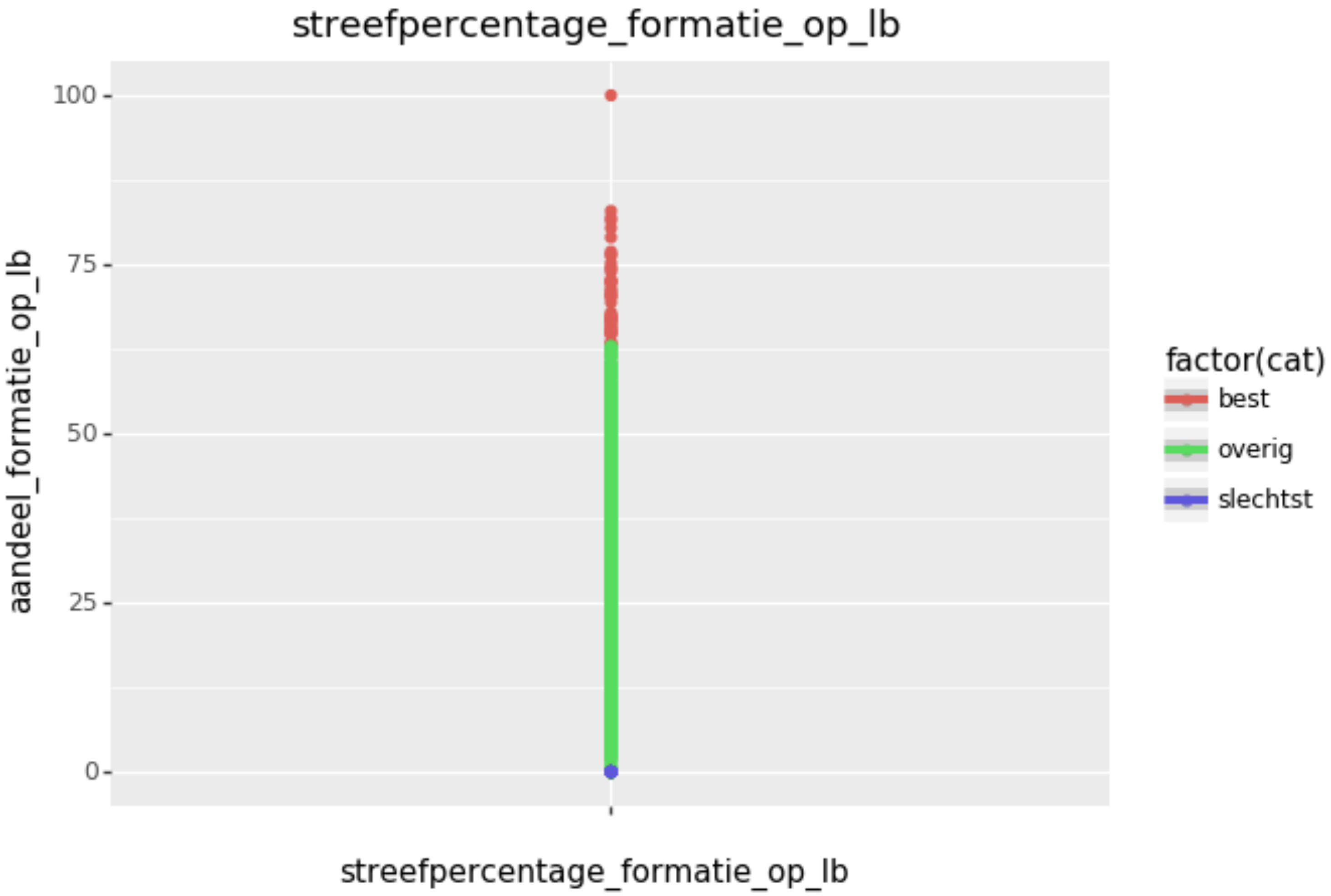
```
<ggplot: (-9223372036563571915)>
aandeel_formatie_op_overig
```



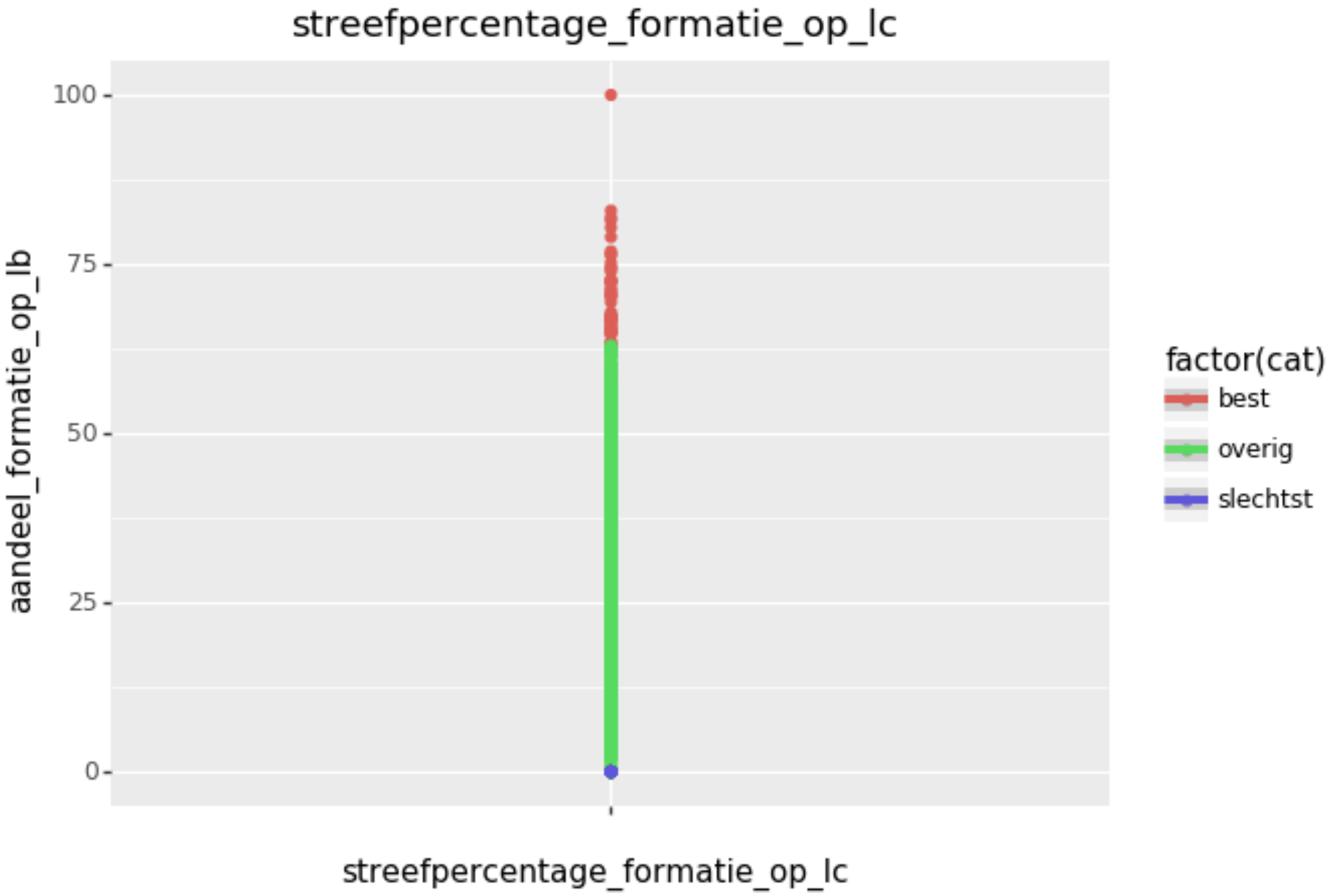
```
<ggplot: (-9223372036561481991)>
streefpercentage_formatie_op_la
```



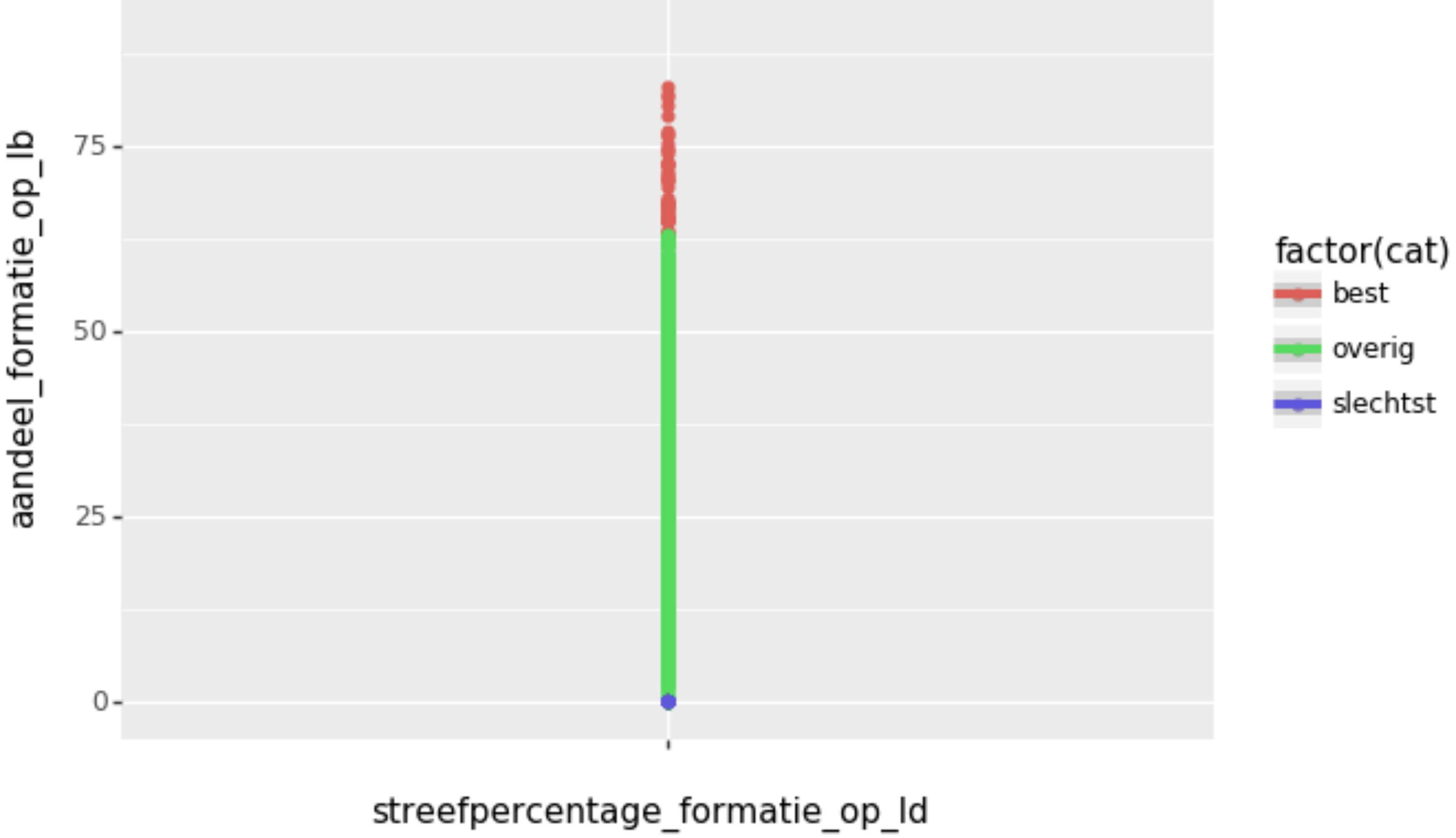
```
<ggplot: (292561279)>
streefpercentage_formatie_op_lb
```



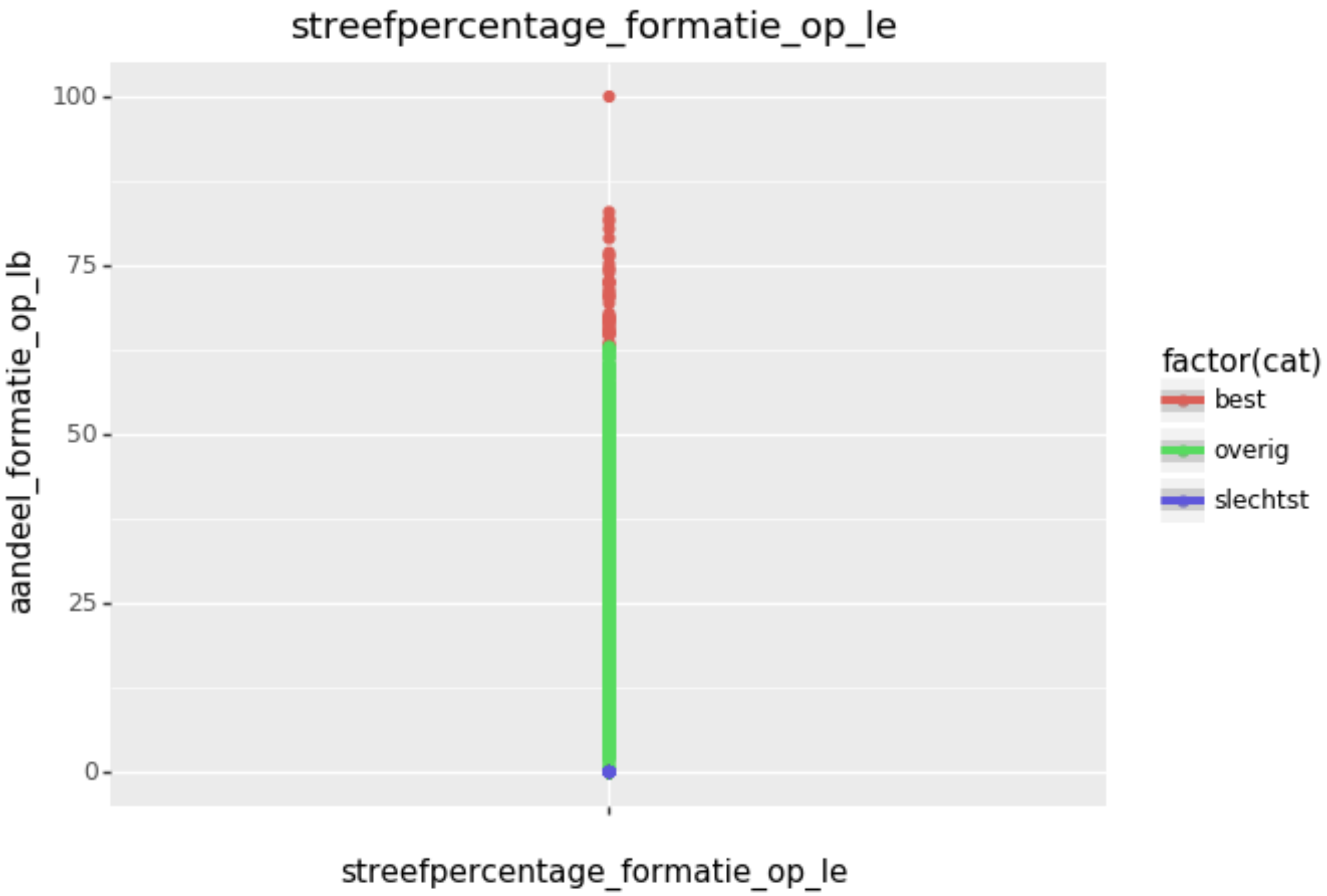
```
<ggplot: (292561279)>
streefpercentage_formatie_op_lc
```



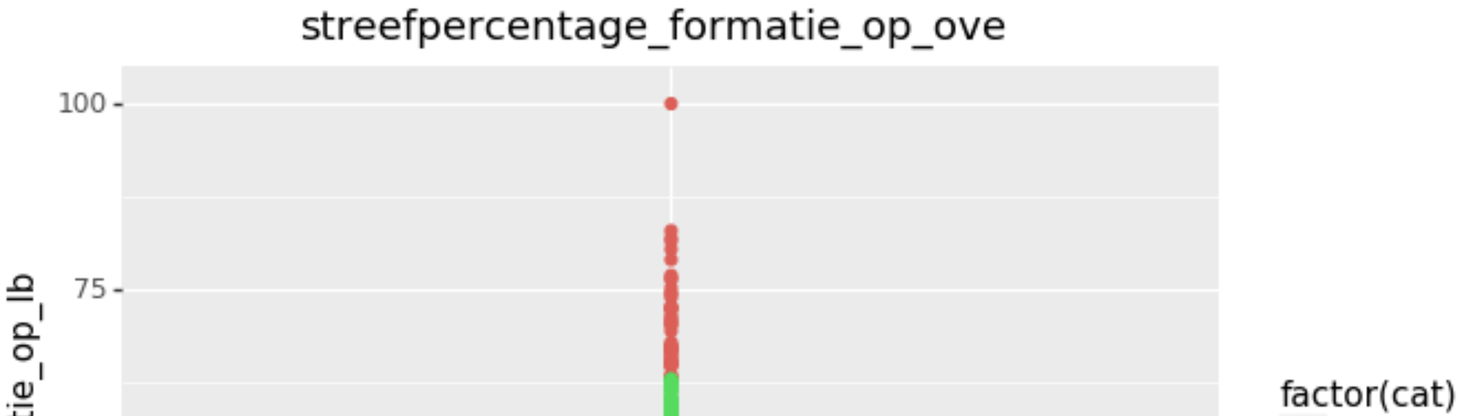
```
<ggplot: (291832896)>
streefpercentage_formatie_op_ld
```

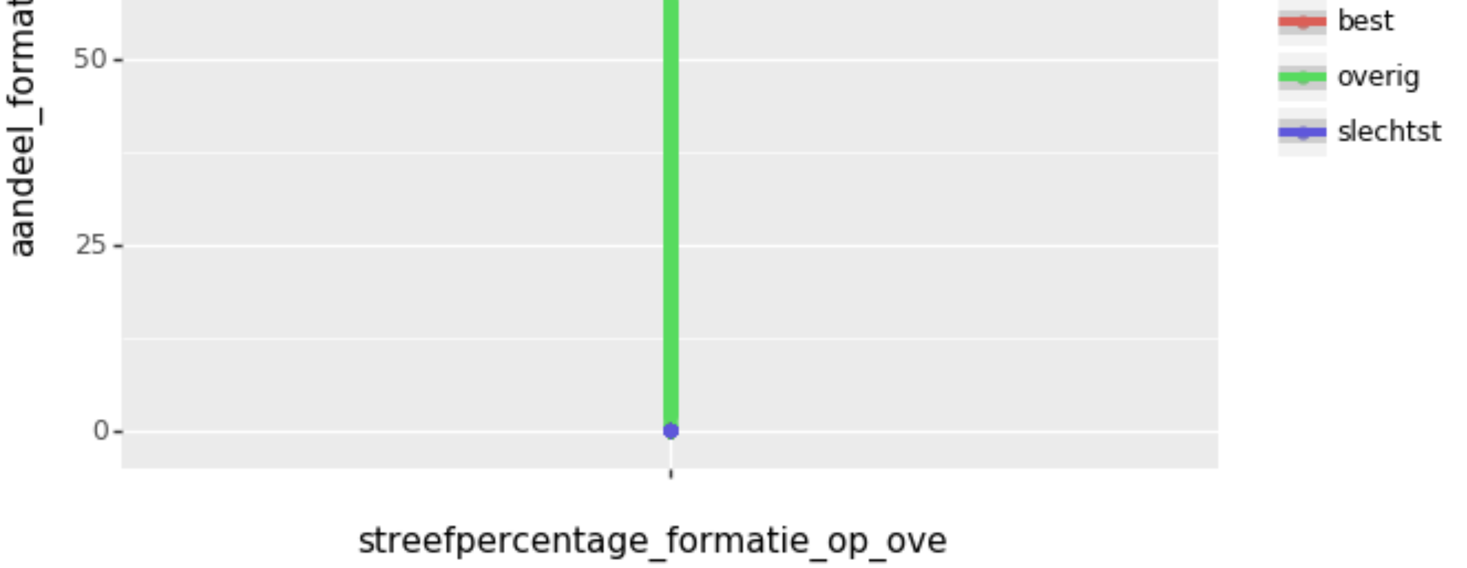


```
<ggplot: (291833001)>
streefpercentage_formatie_op_le
```

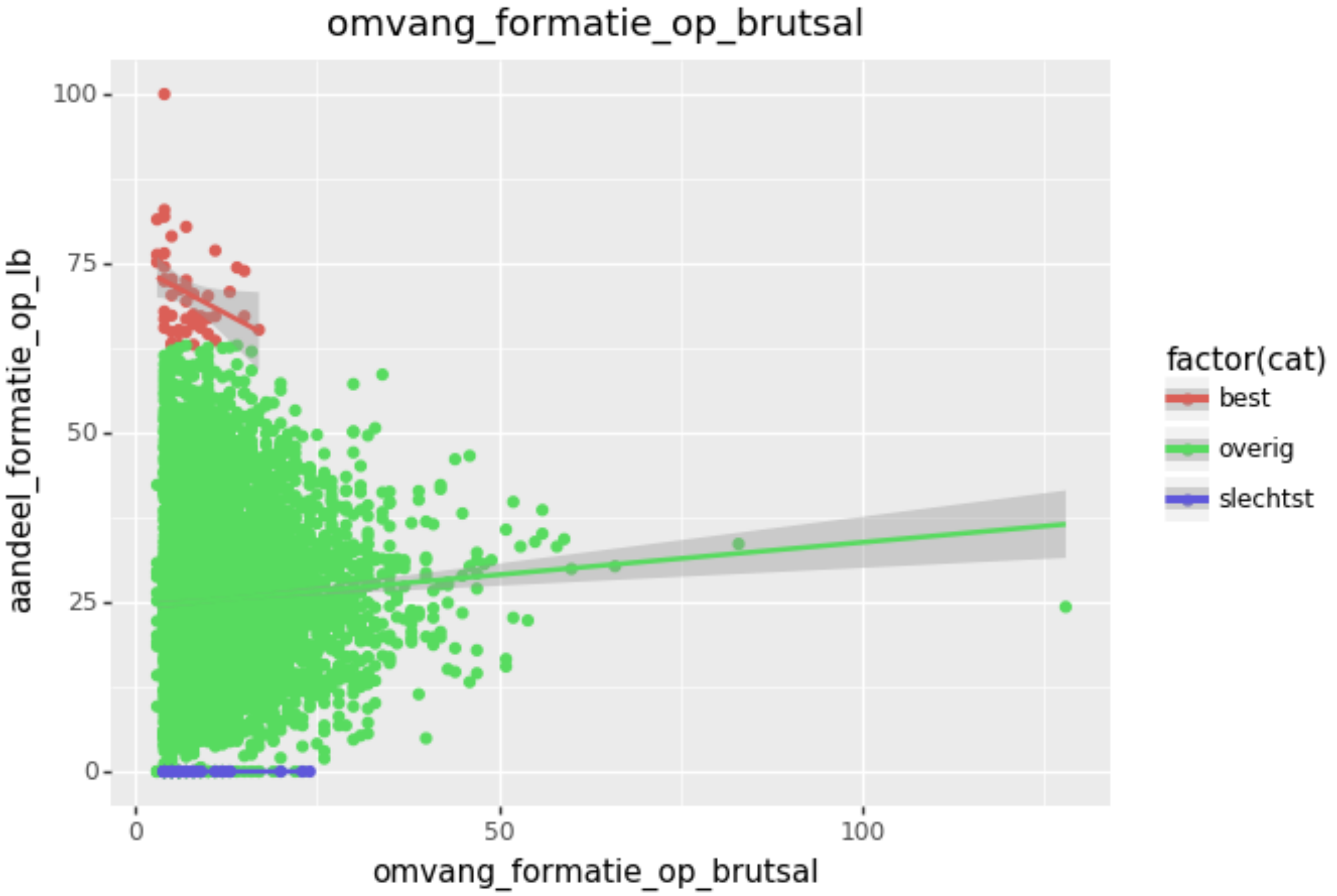


```
<ggplot: (-9223372036576248060)>
streefpercentage_formatie_op_ove
```

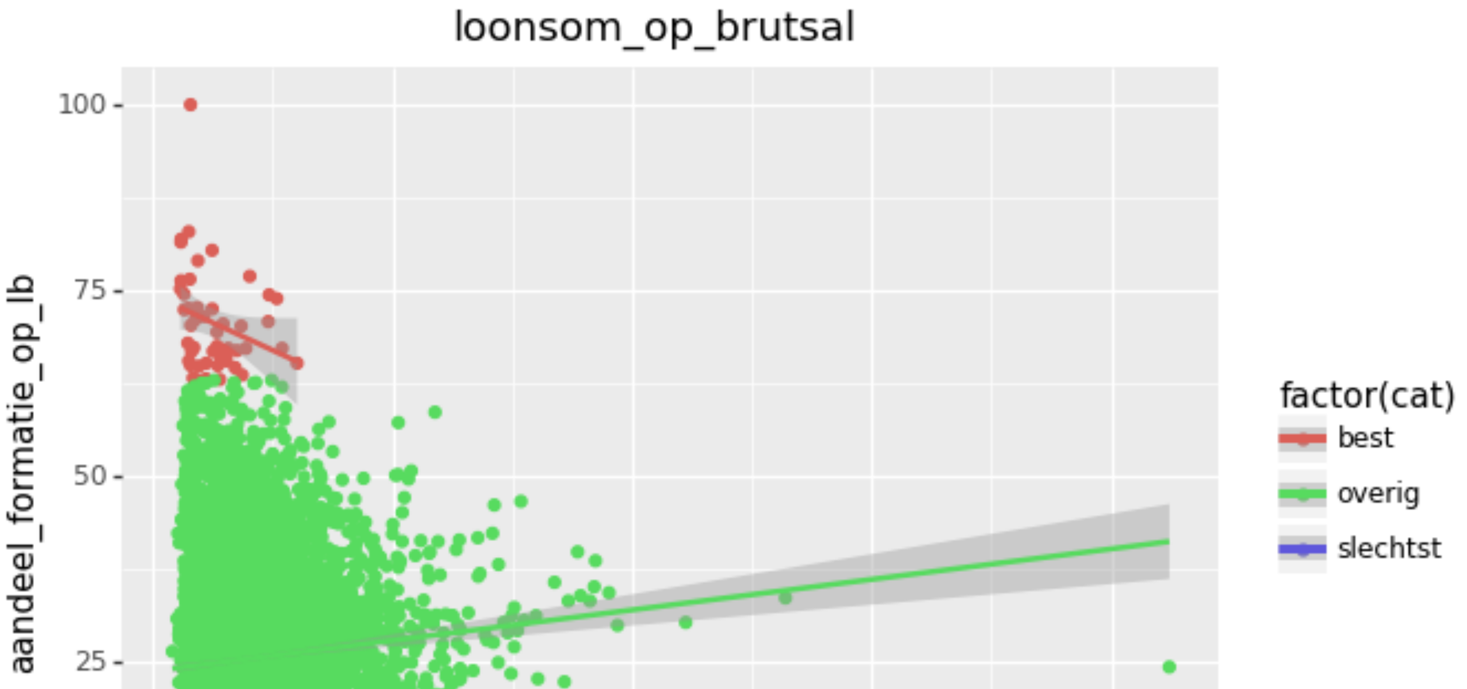


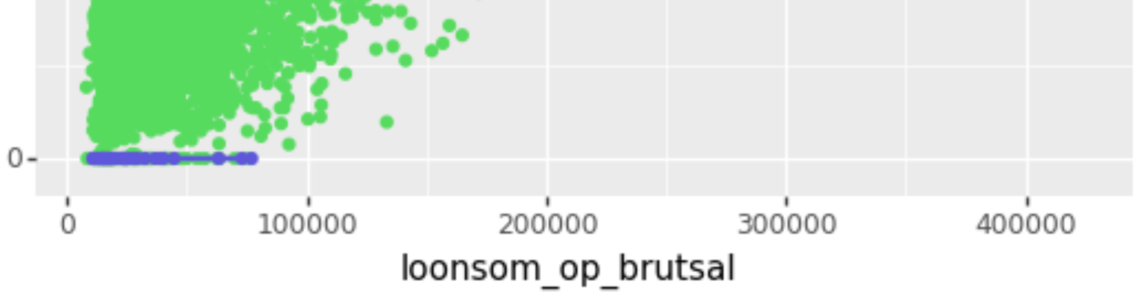


```
<ggplot: (278535579)>
omvang_formatie_op_brutsal
```

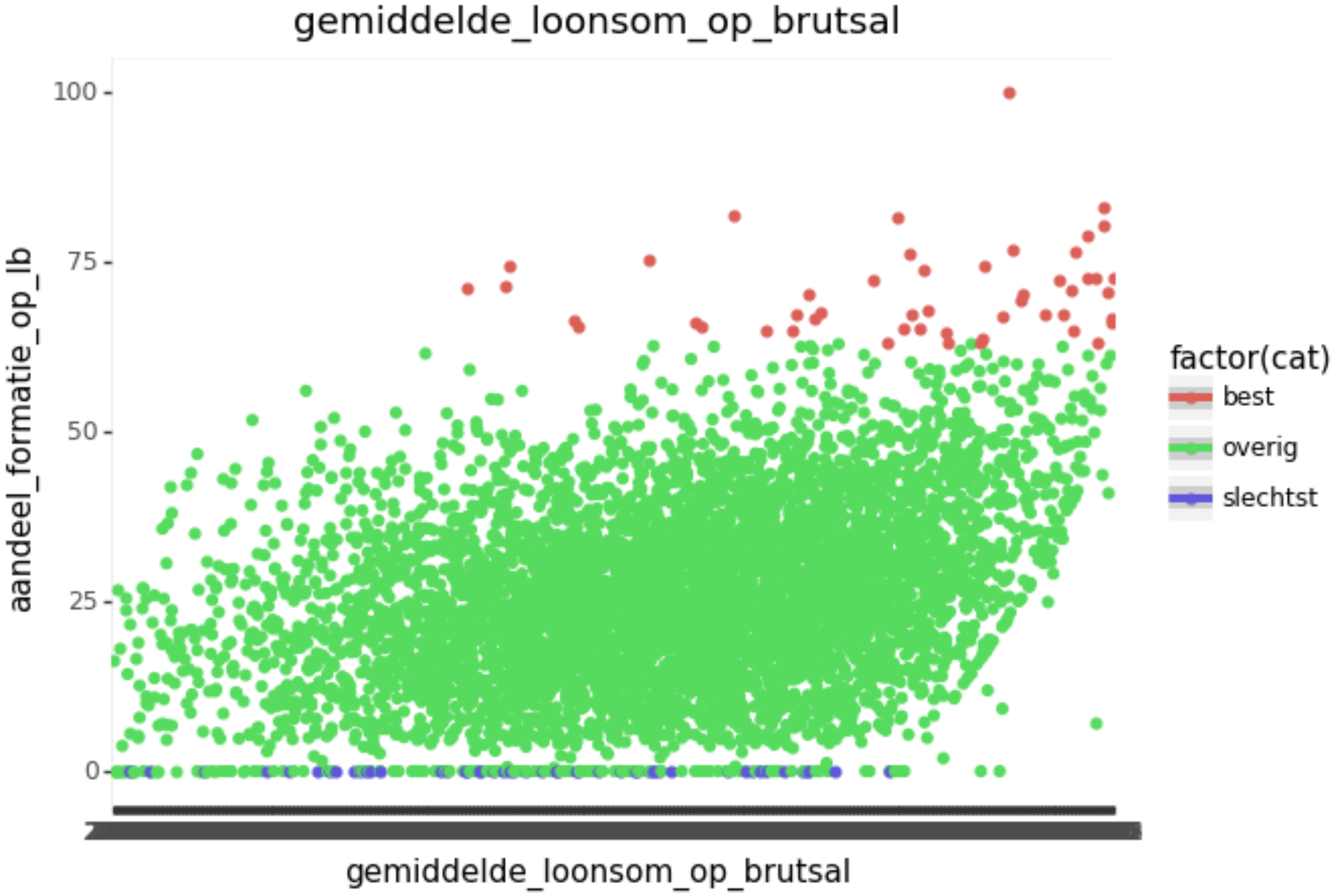


```
<ggplot: (-9223372036562935927)>
loonsom_op_brutsal
```

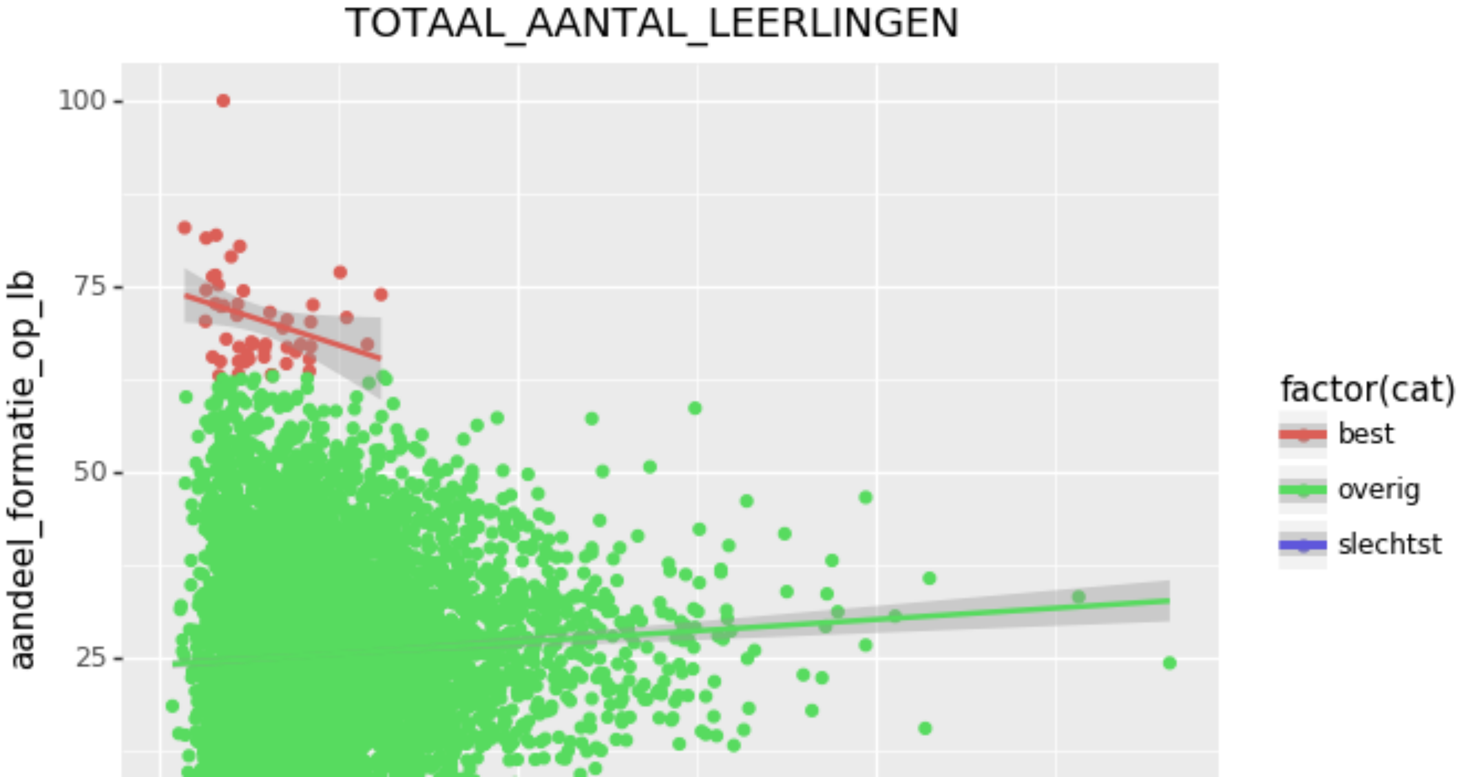


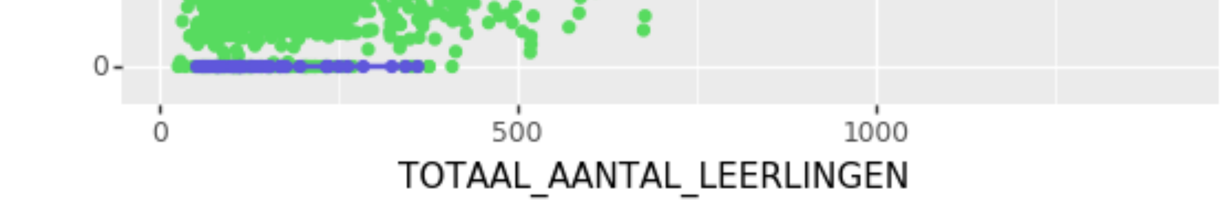


```
<ggplot: (286560106)>
gemiddelde_loonsom_op_brutsal
```

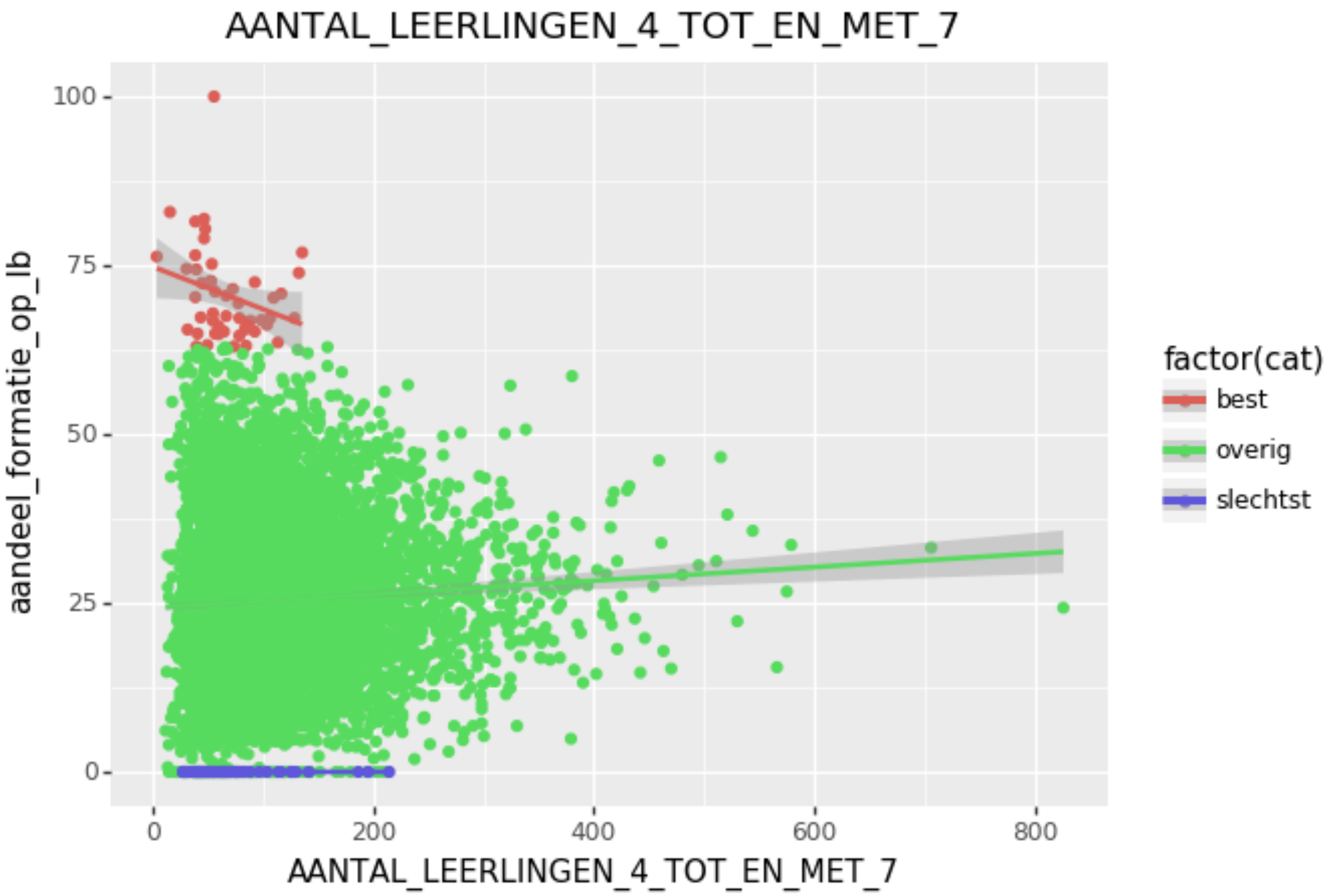


```
<ggplot: (286560106)>
PEILDATUM
NO NON-NA ROWS
BEVOEGD_GEZAG_NUMMER
NO NON-NA ROWS
TOTAAL_AANTAL_LEERLINGEN
```

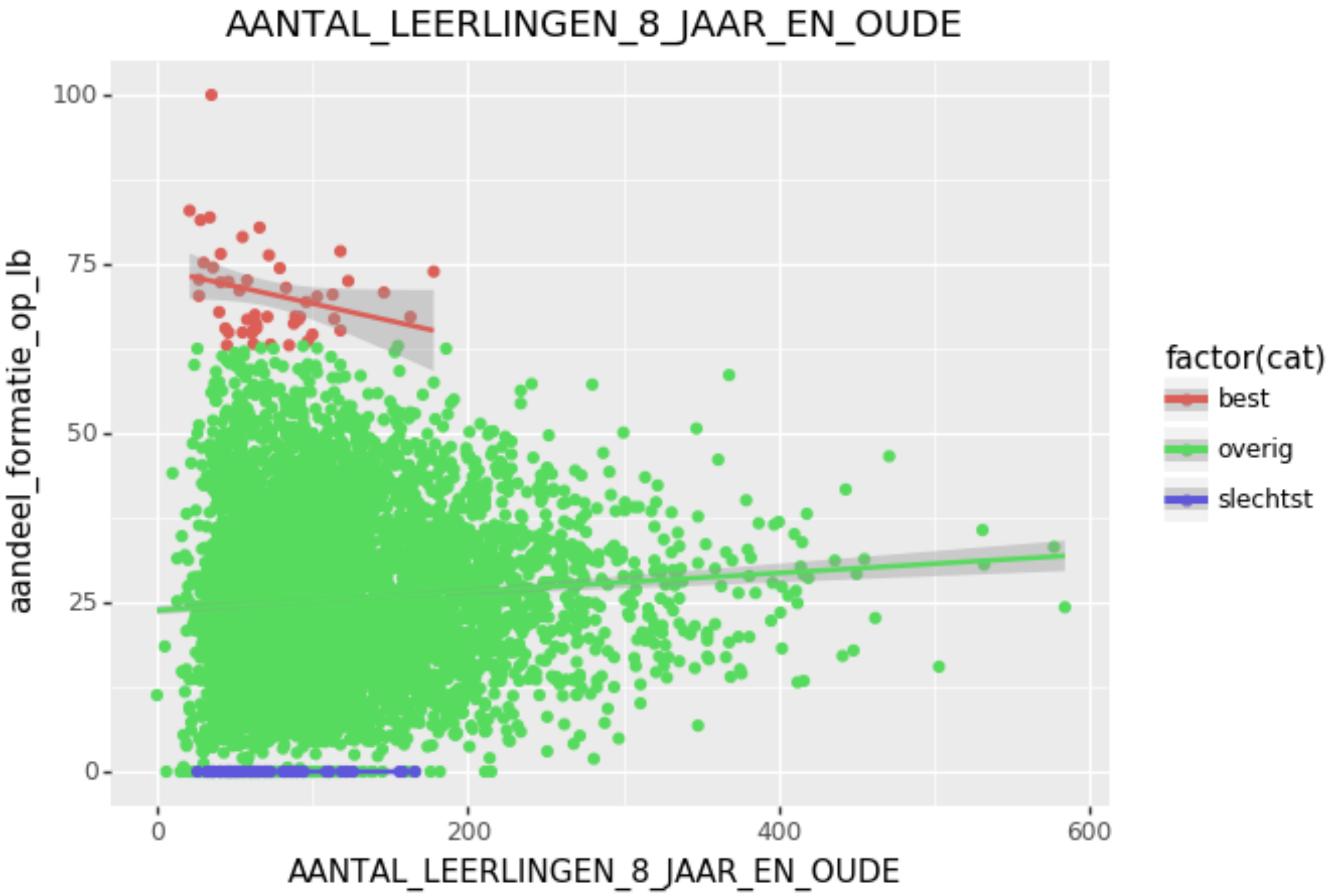




```
<ggplot: (287225096)>
AANTAL_LEERLINGEN_4_TOT_EN_MET_7
```

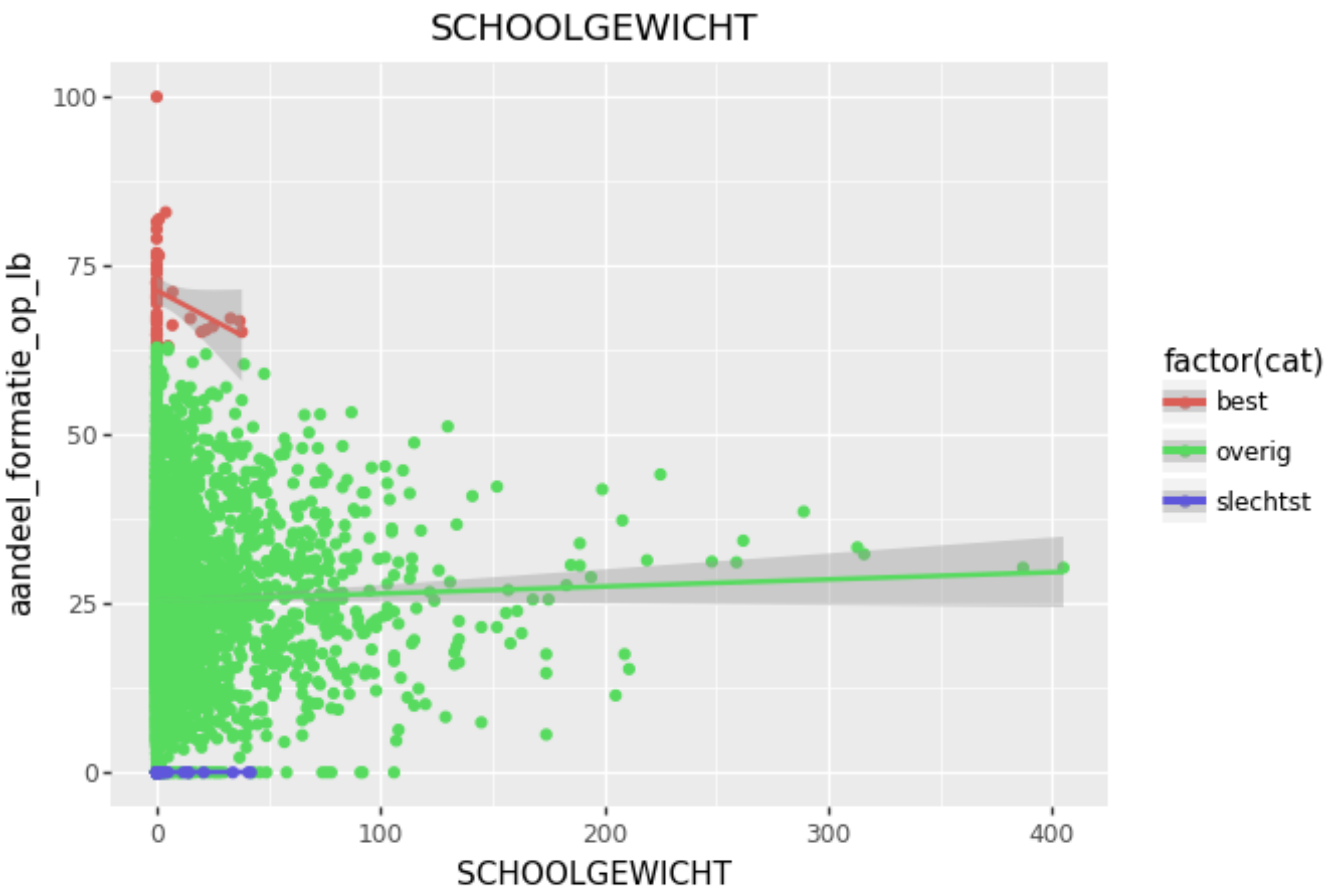


```
<ggplot: (291447002)>
AANTAL_LEERLINGEN_8_JAAR_EN_OUDE
```

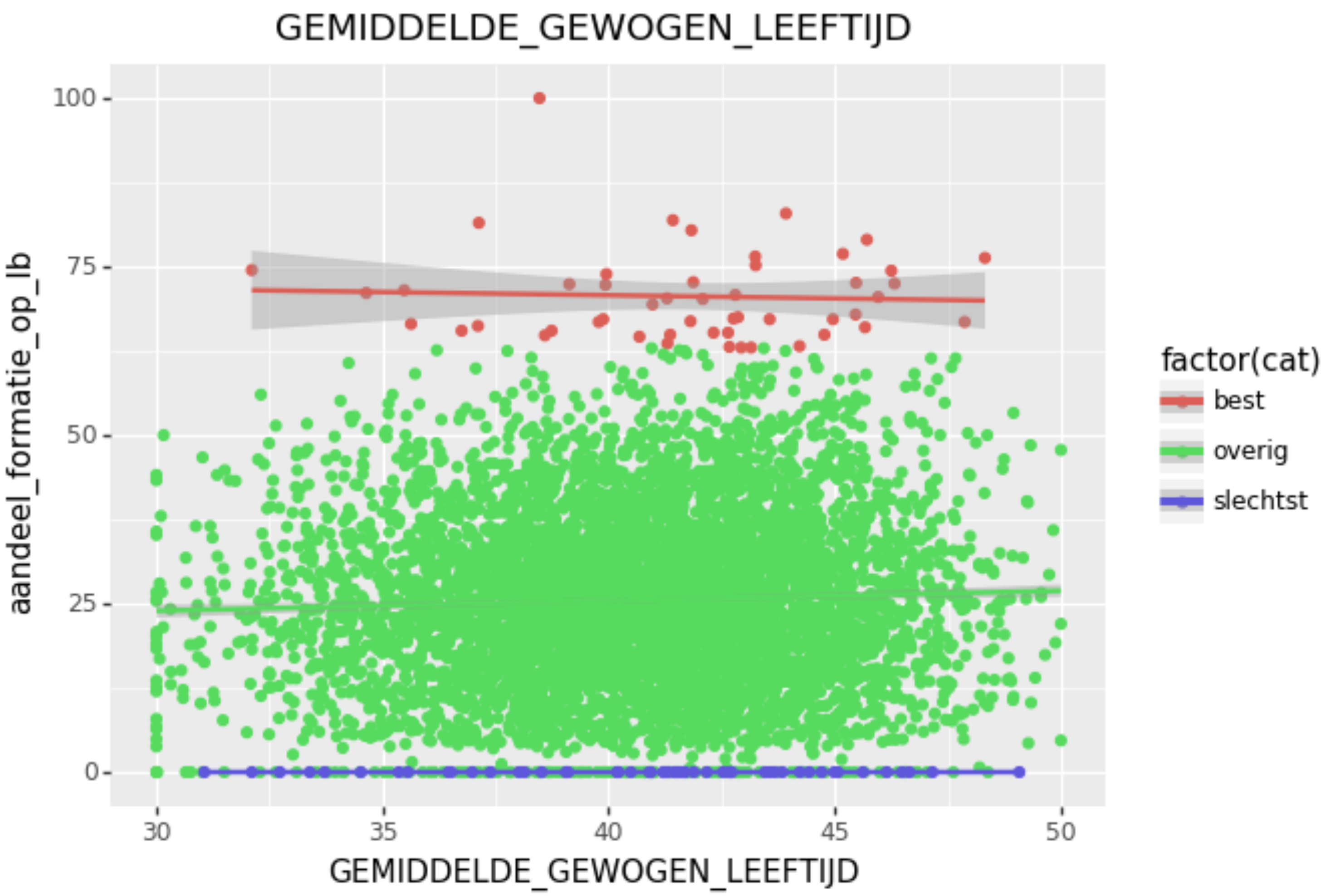


```
<ggplot: (-9223372036561558328)>
```

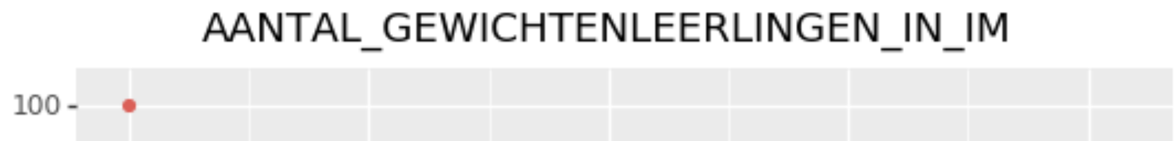
```
<ggplot: (-9223372036561558328)>
SCHOOLGEWICHT
```

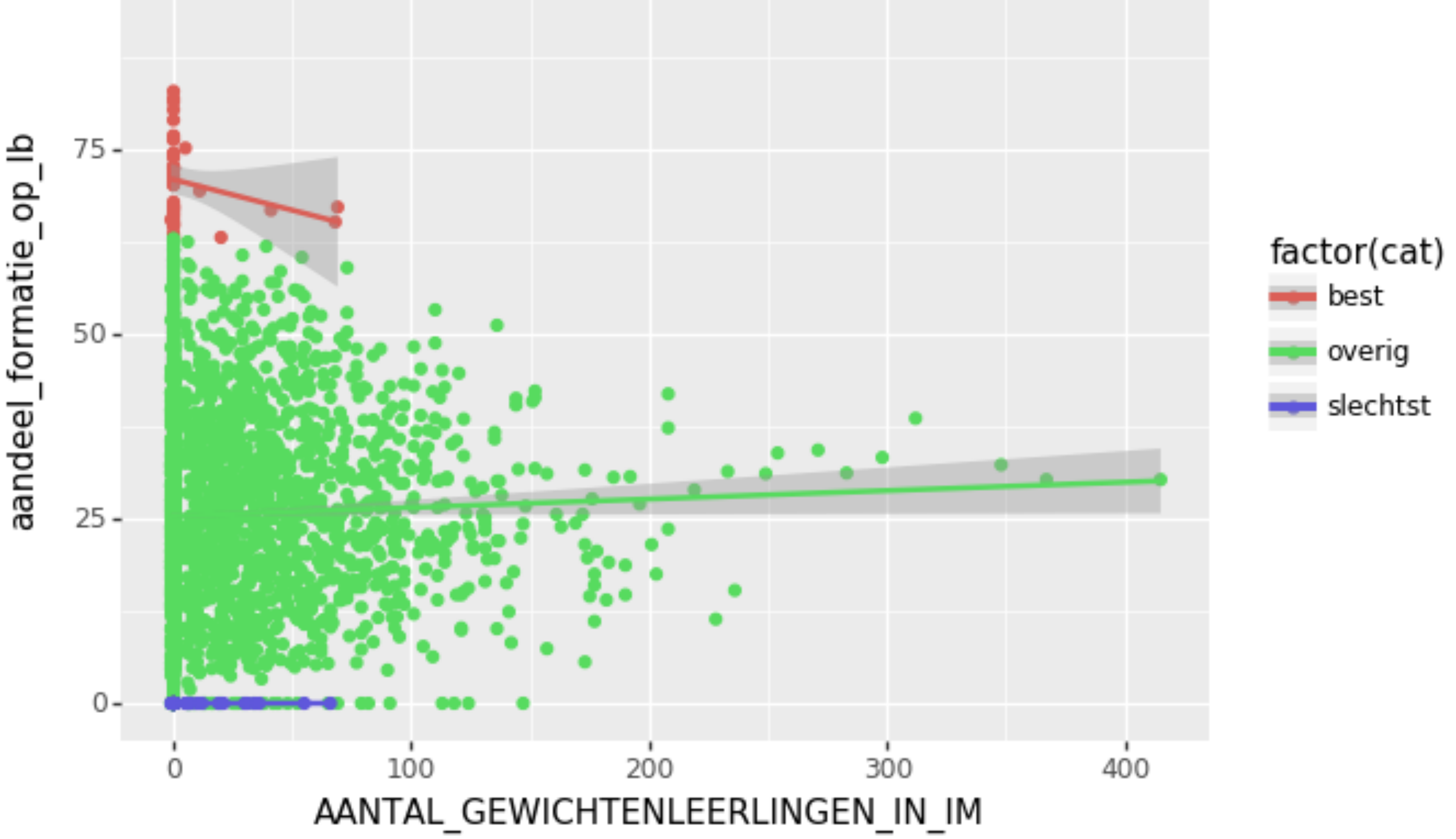


```
<ggplot: (-9223372036561455030)>
GEMIDDELDE_GEWOGEN_LEEFTIJD
```

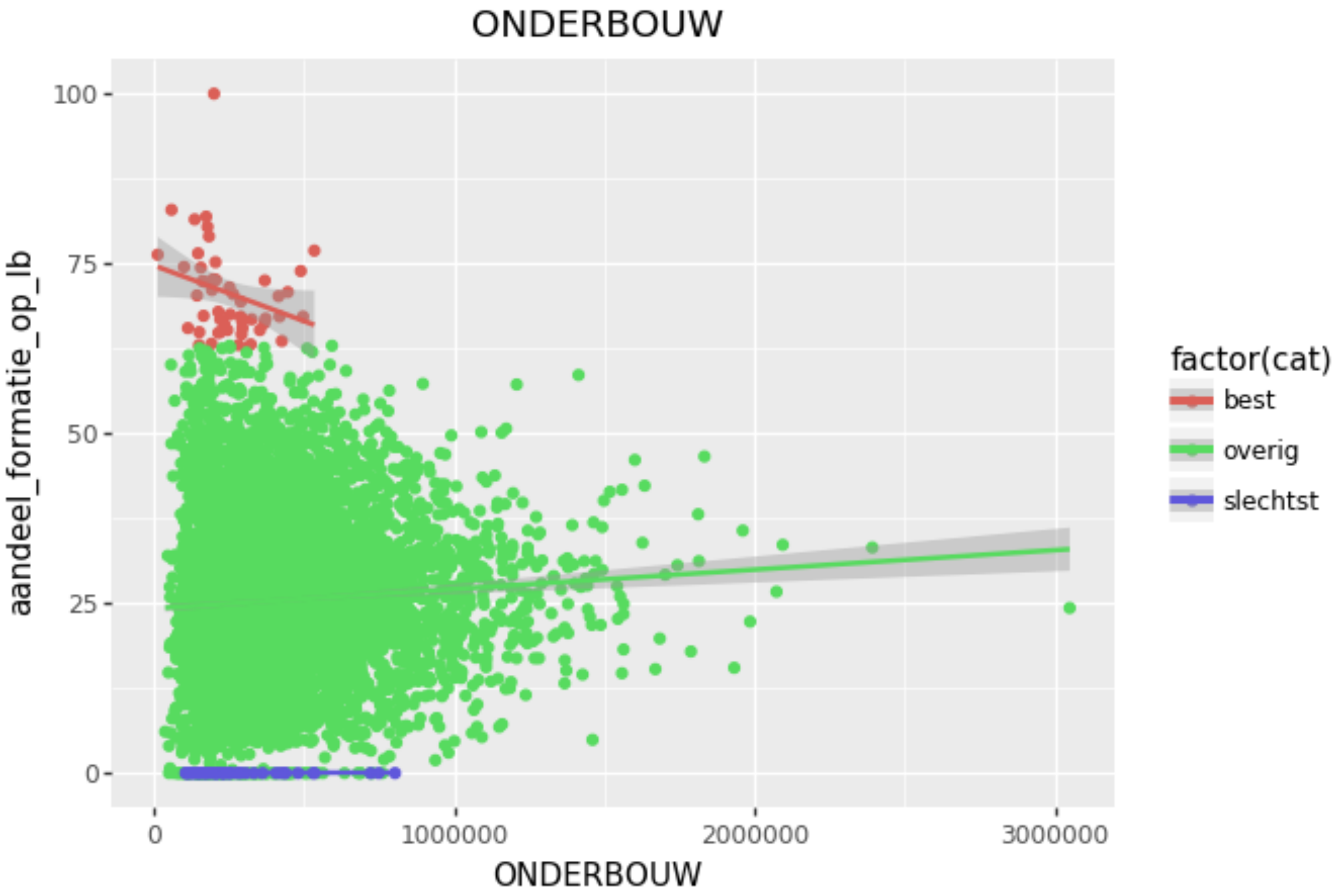


```
<ggplot: (293320768)>
AANTAL_GEWICHTENLEERLINGEN_IN_IM
```

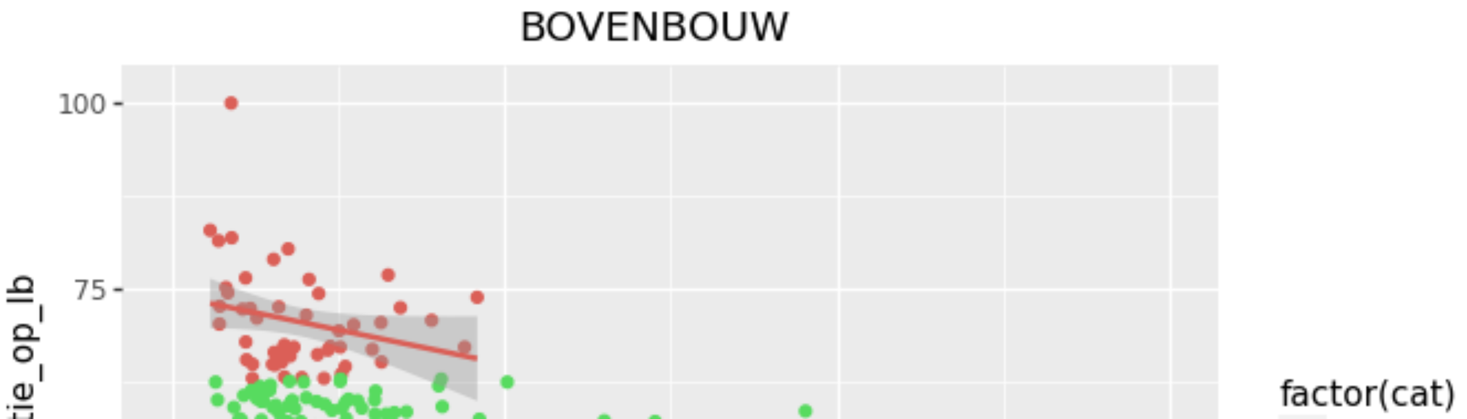


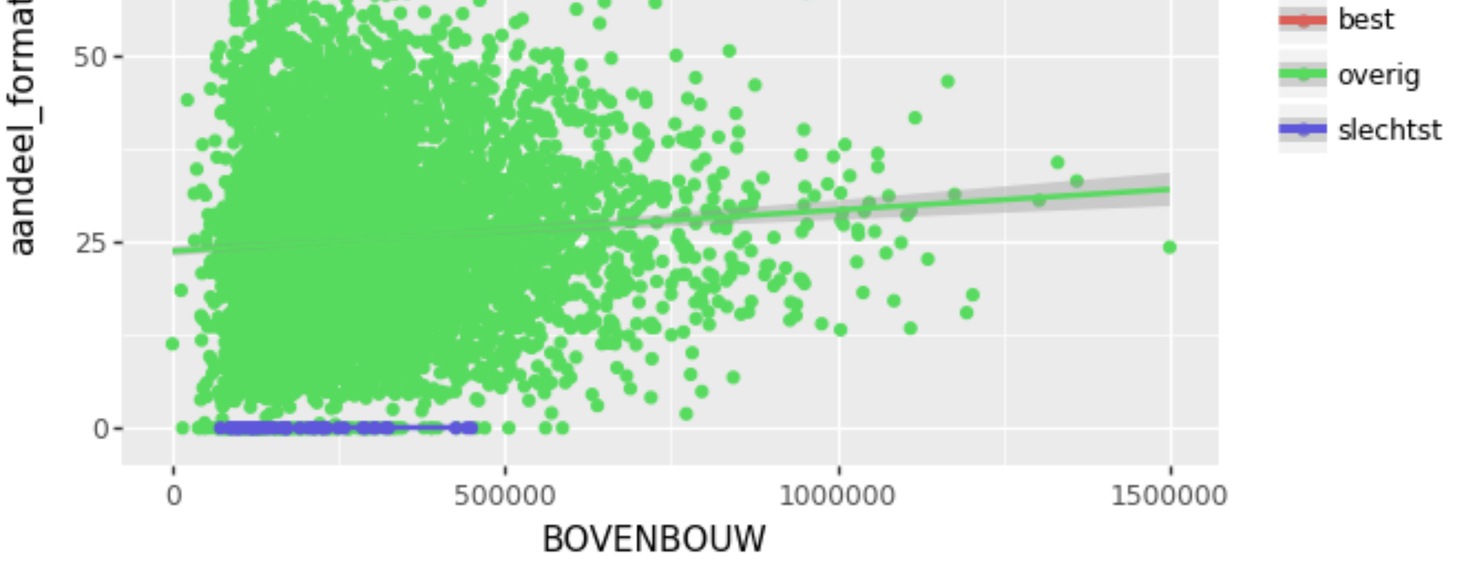


```
<ggplot: (290426737)>
ONDERBOUW
```

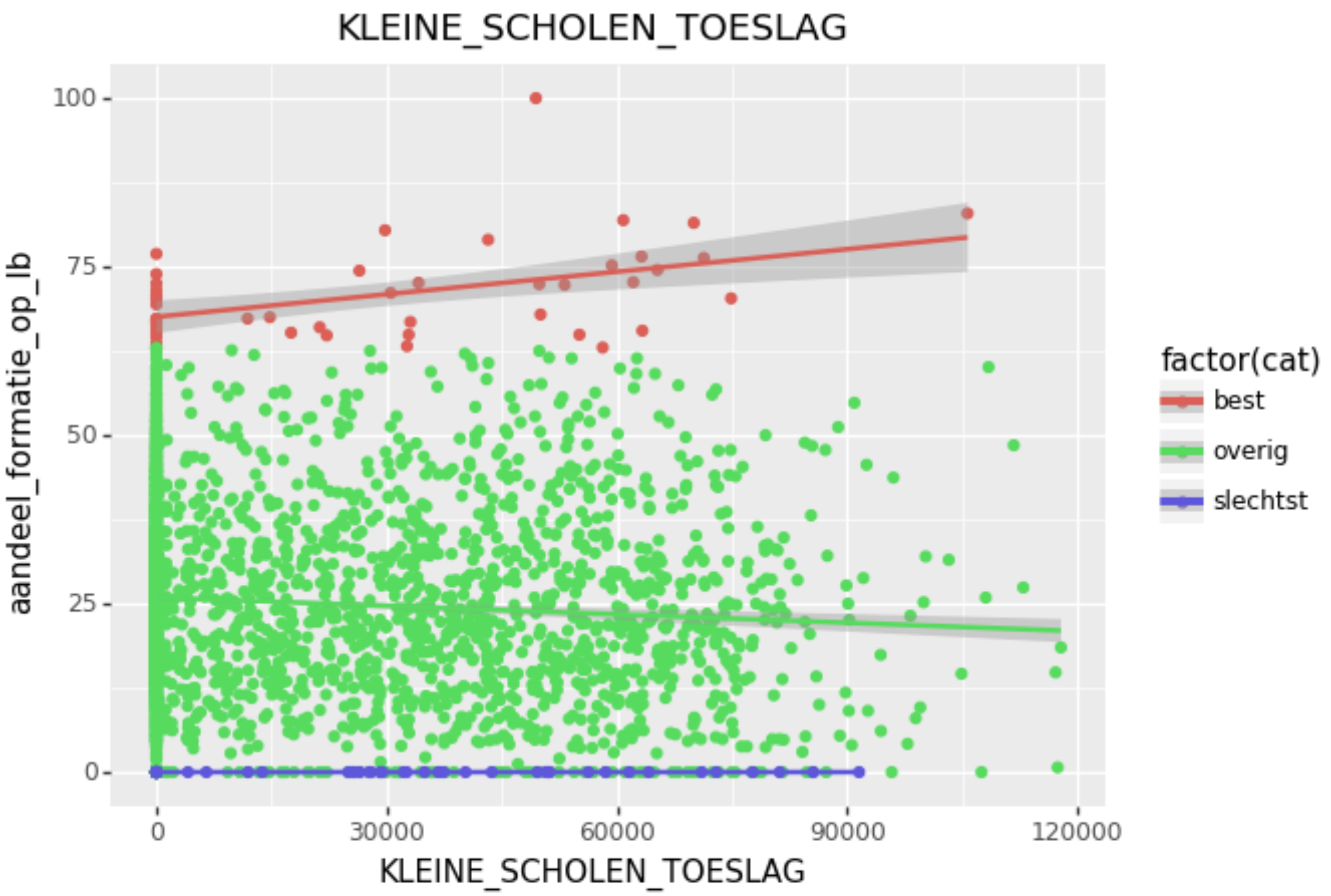


```
<ggplot: (-9223372036569093702)>
BOVENBOUW
```

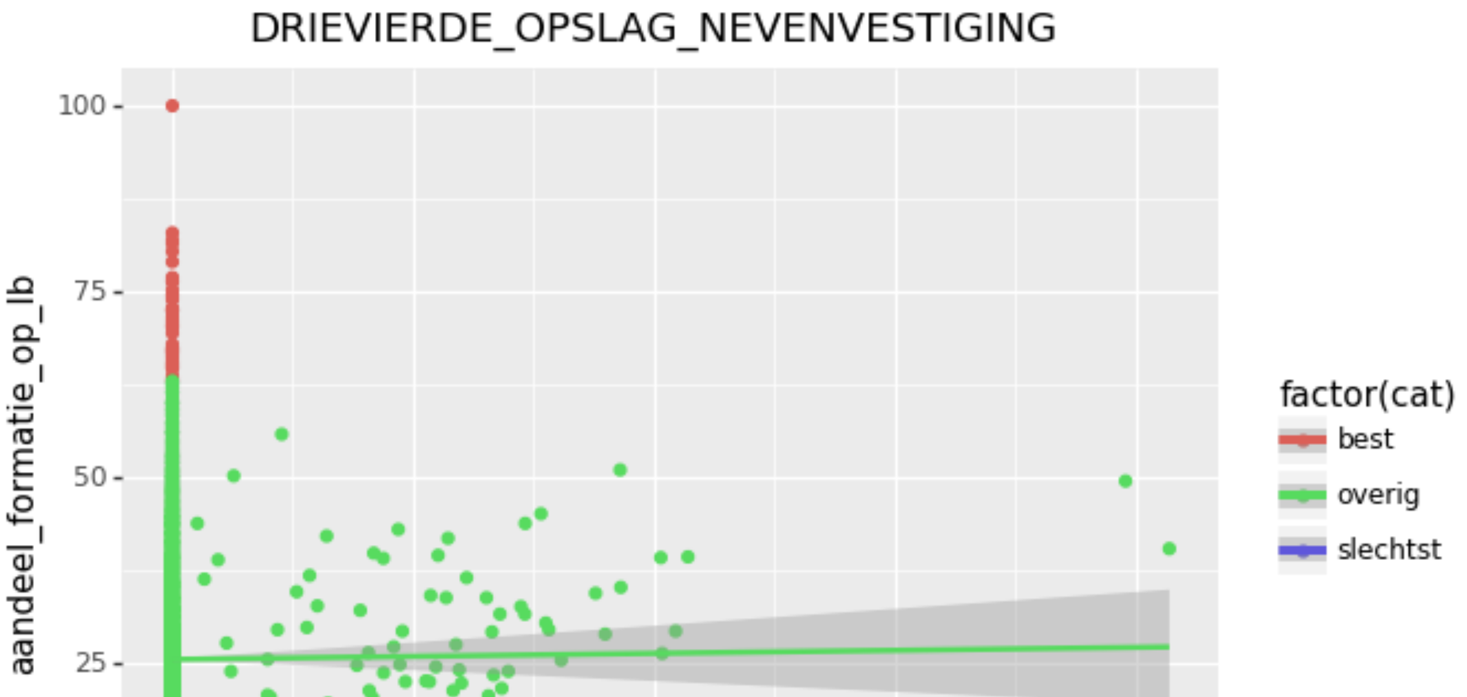


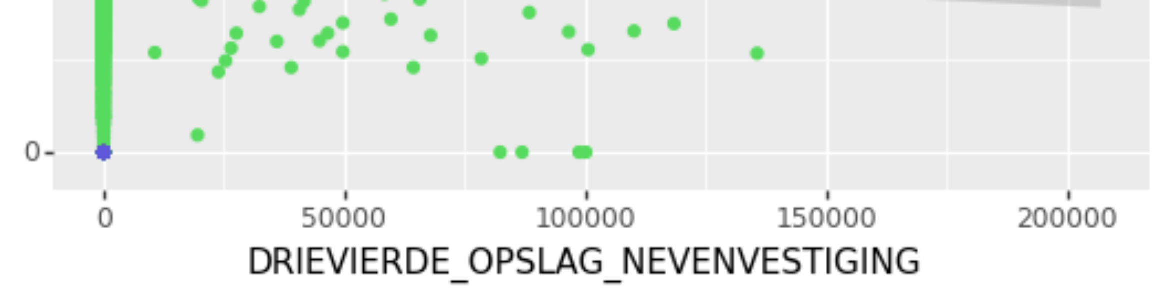


```
<ggplot: (289763055)>
KLEINE_SCHOLEN_TOESLAG
```

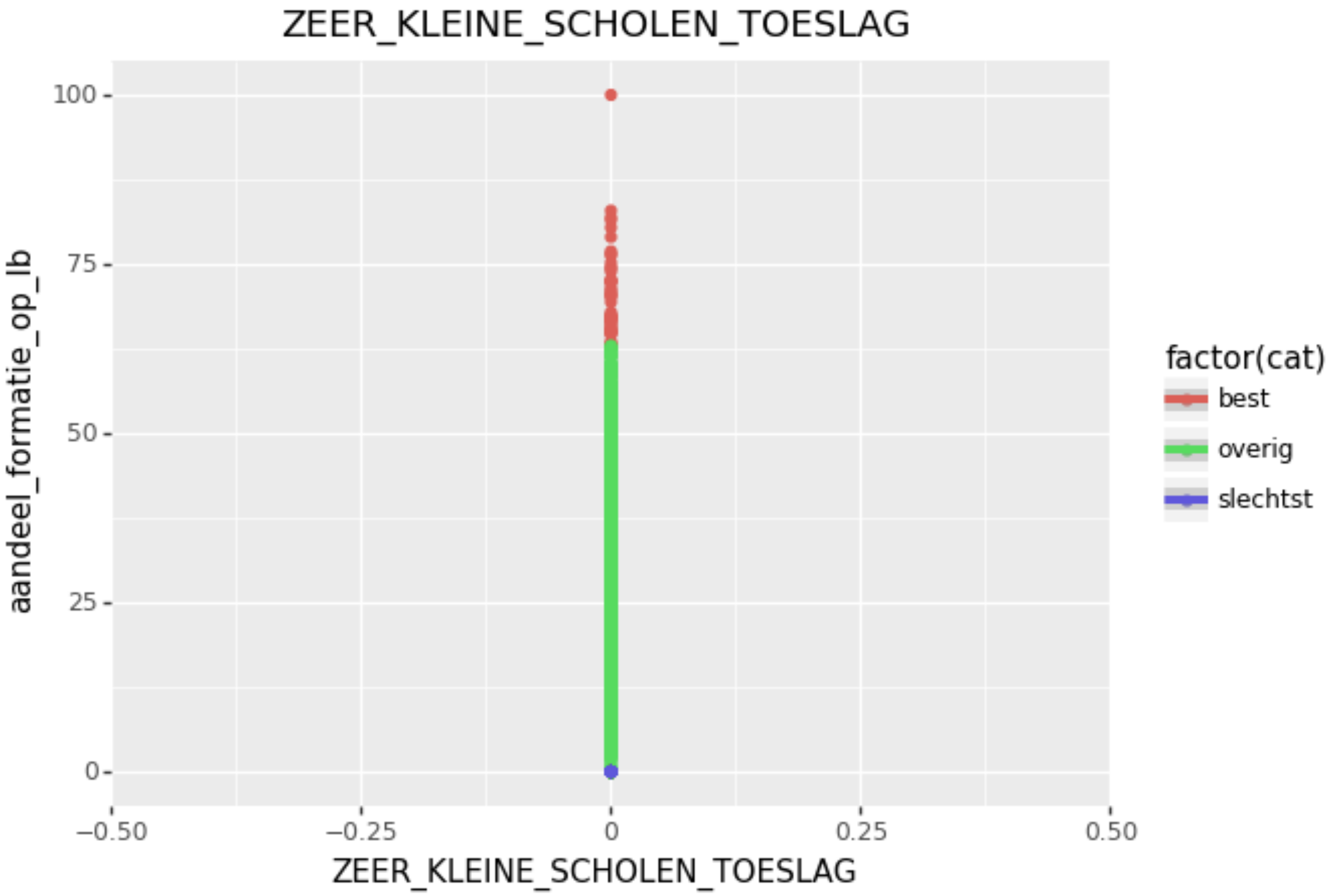


```
<ggplot: (-9223372036564249063)>
DRIEVIERDE_OPSLAG_NEVENVESTIGING
```

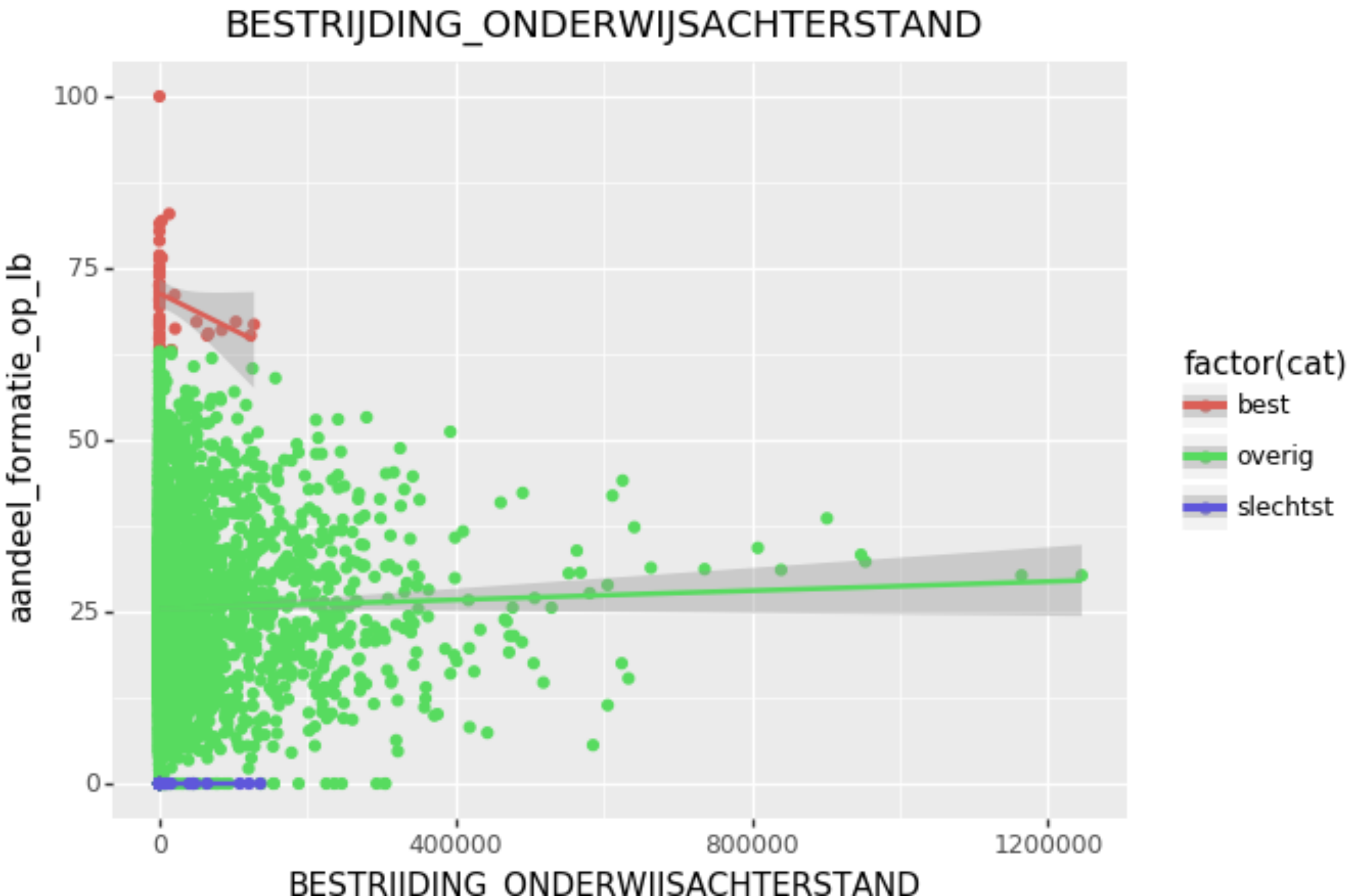




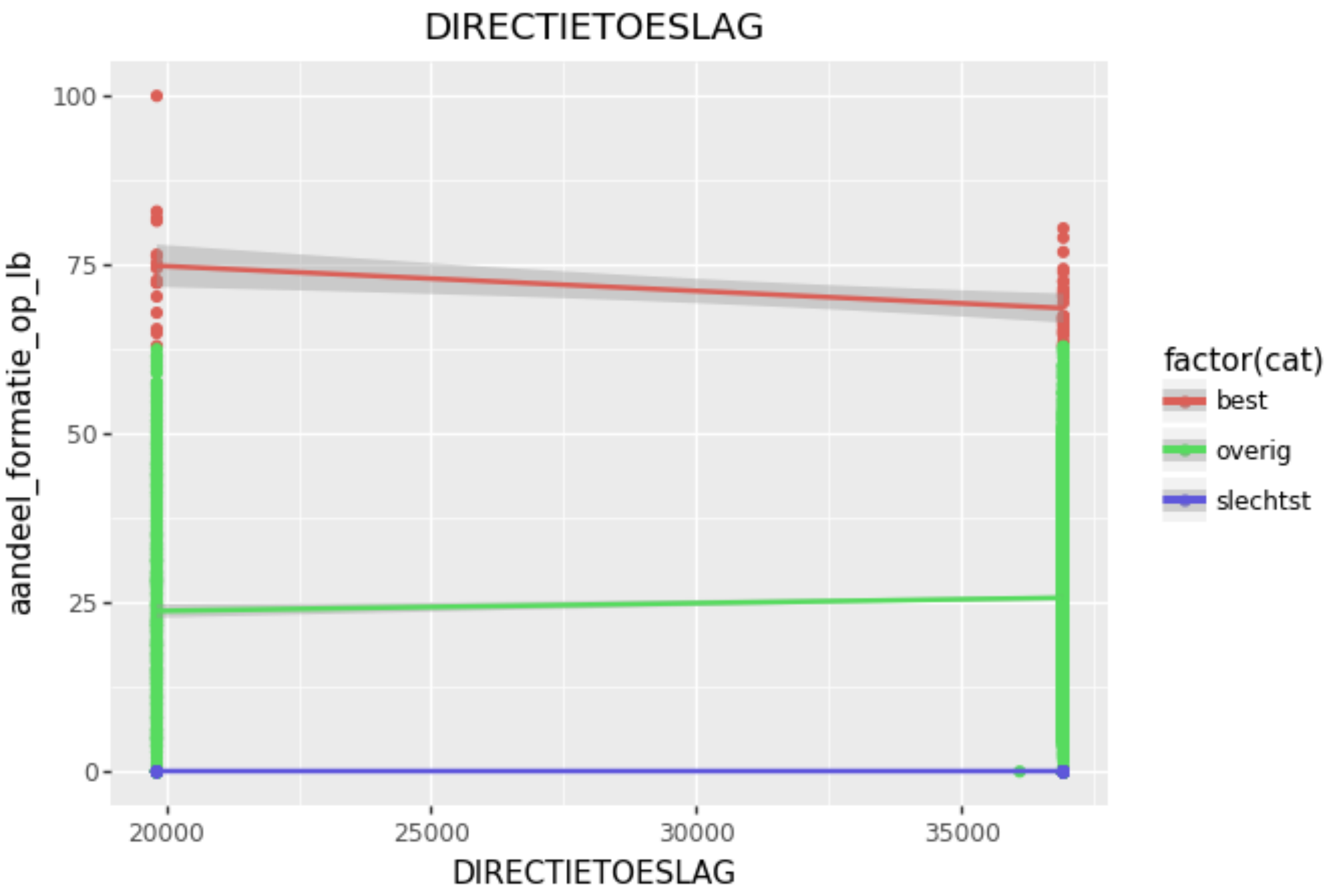
```
<ggplot: (290526728)>
ZEER_KLEINE_SCHOLEN_TOESLAG
```



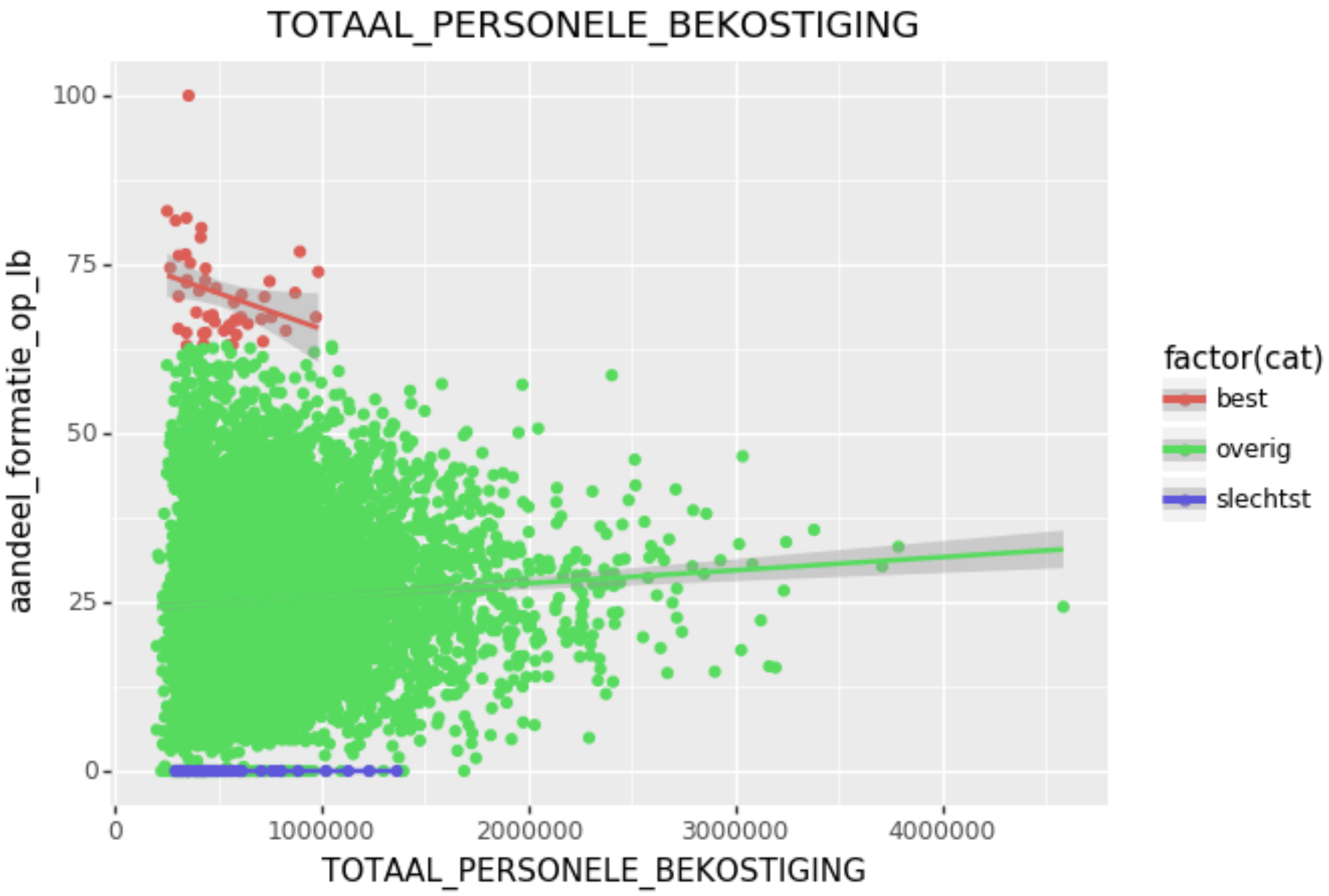
```
<ggplot: (291207246)>
BESTRIJDING_ONDERWIJSACHTERSTAND
```



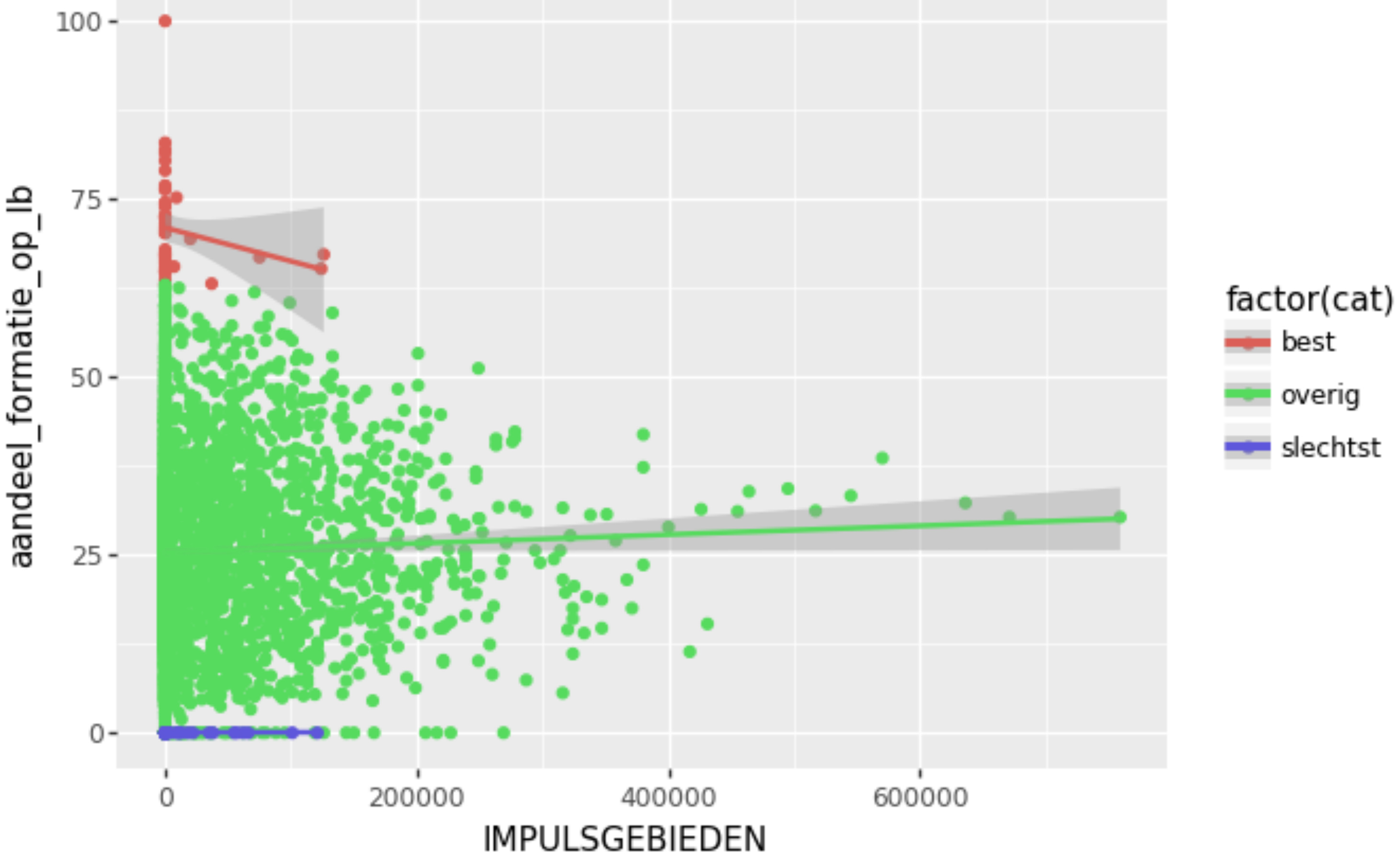
```
<ggplot: (292067897)>
DIRECTIETOESLAG
```



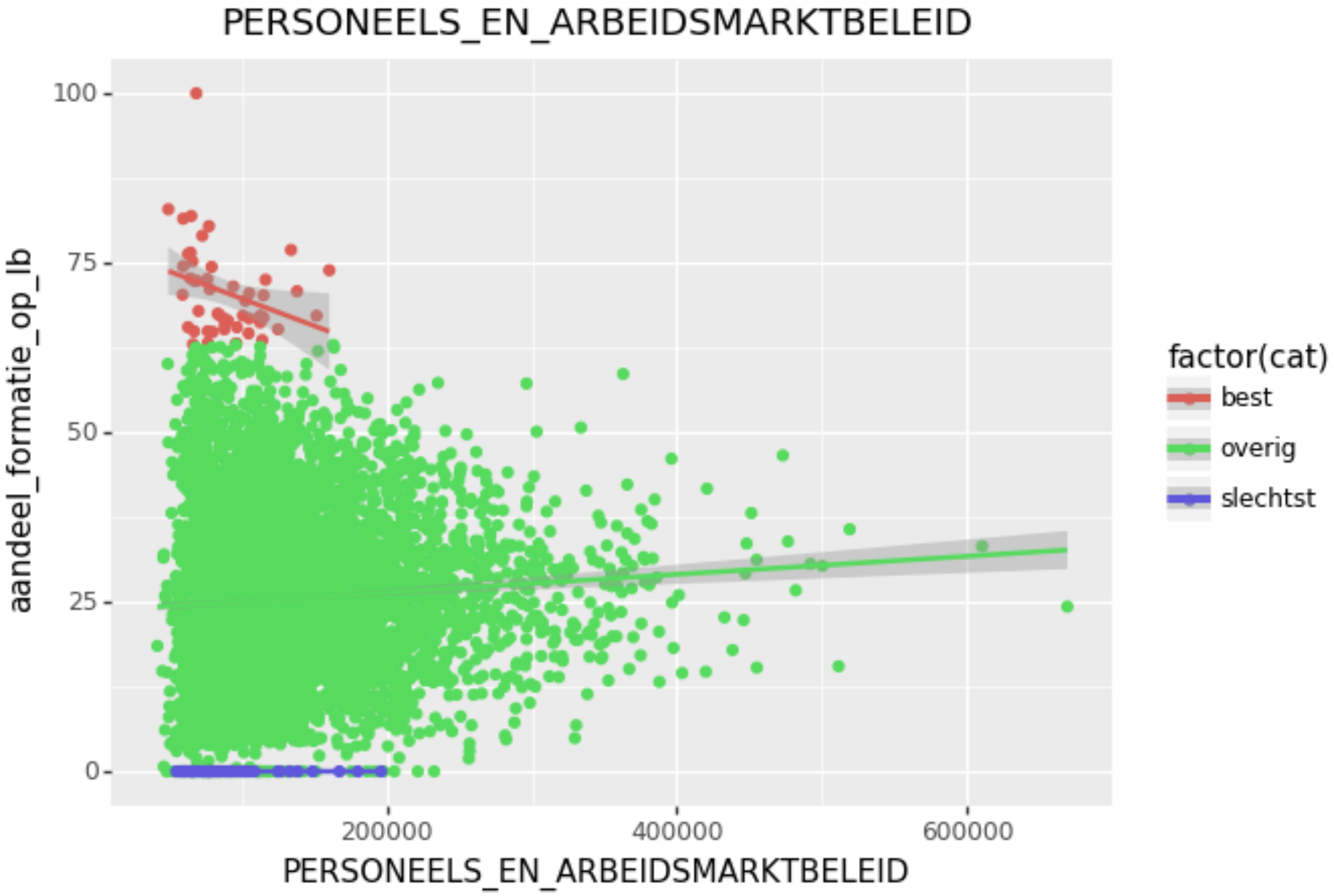
```
<ggplot: (-9223372036562326710)>
TOTAAL_PERSONELE_BEKOSTIGING
```



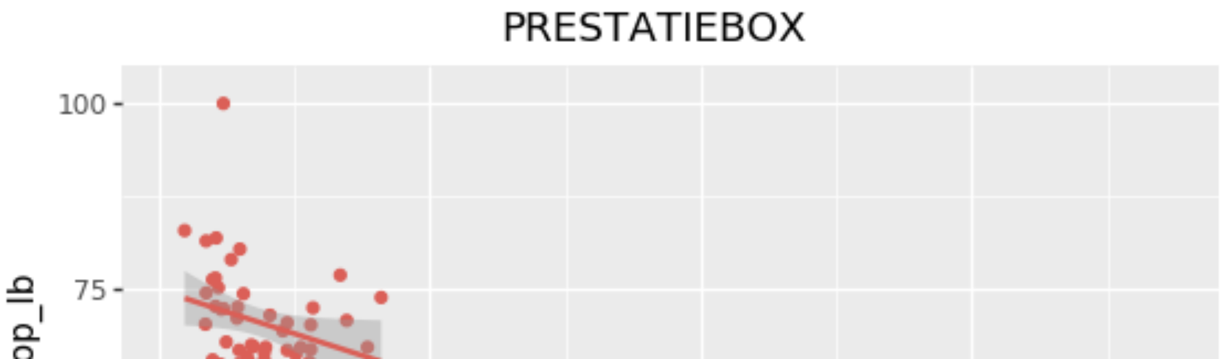
```
<ggplot: (-9223372036561729287)>
IMPULSGEBIEDEN
```

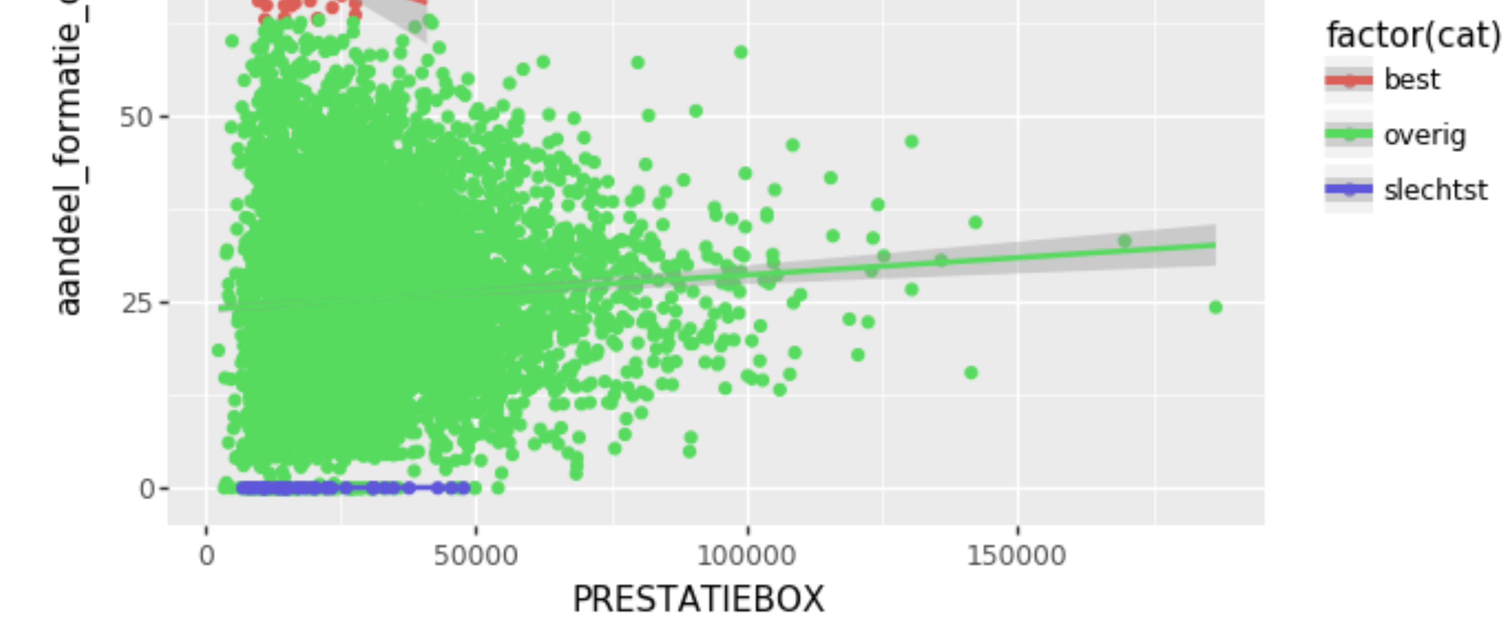


```
<ggplot: (-9223372036561729371)>
PERSONEELS_EN_ARBEIDSMARKTBELEID
```

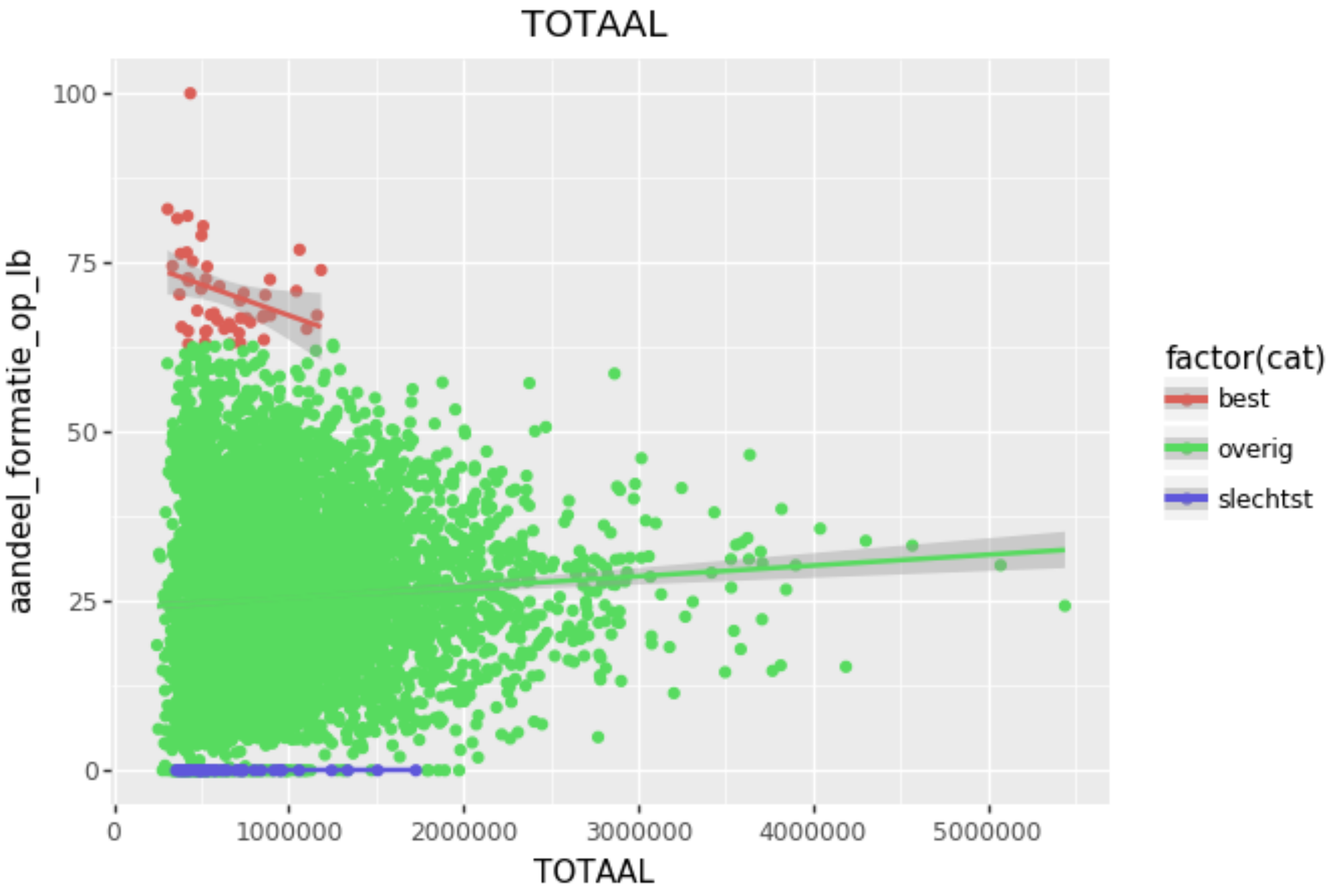


```
<ggplot: (278540580)>
PRESTATIEBOX
```

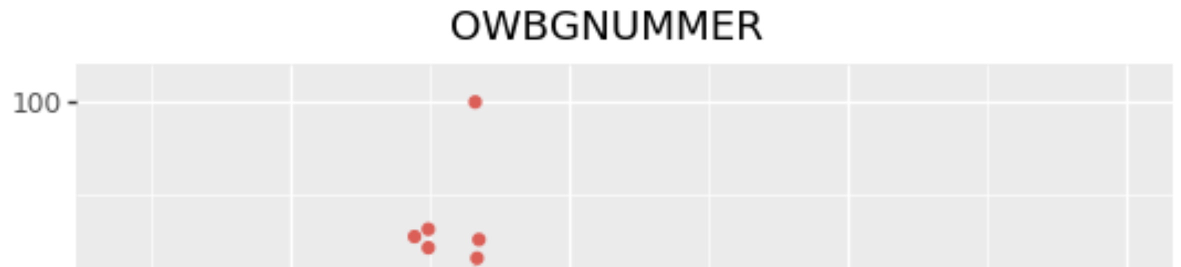


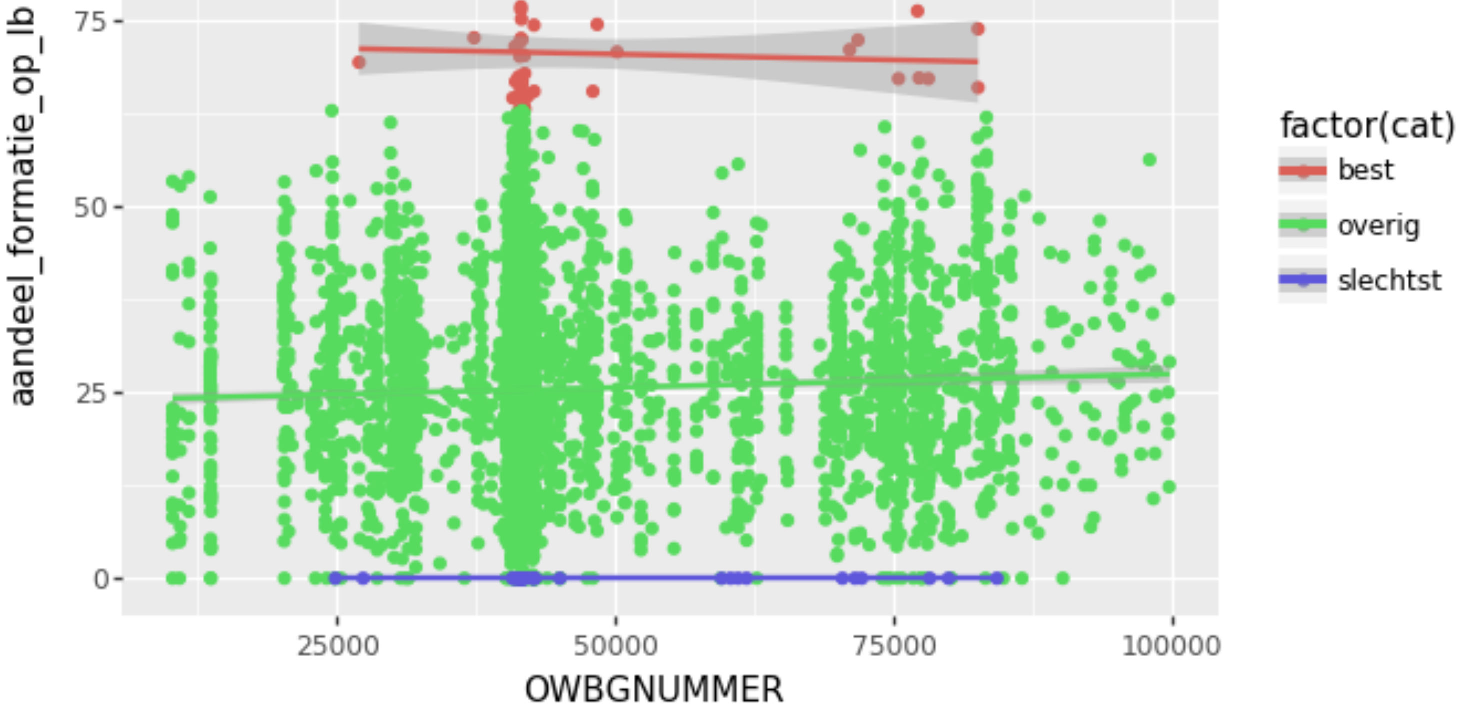


```
<ggplot: (290701537)>
TOTAAL
```

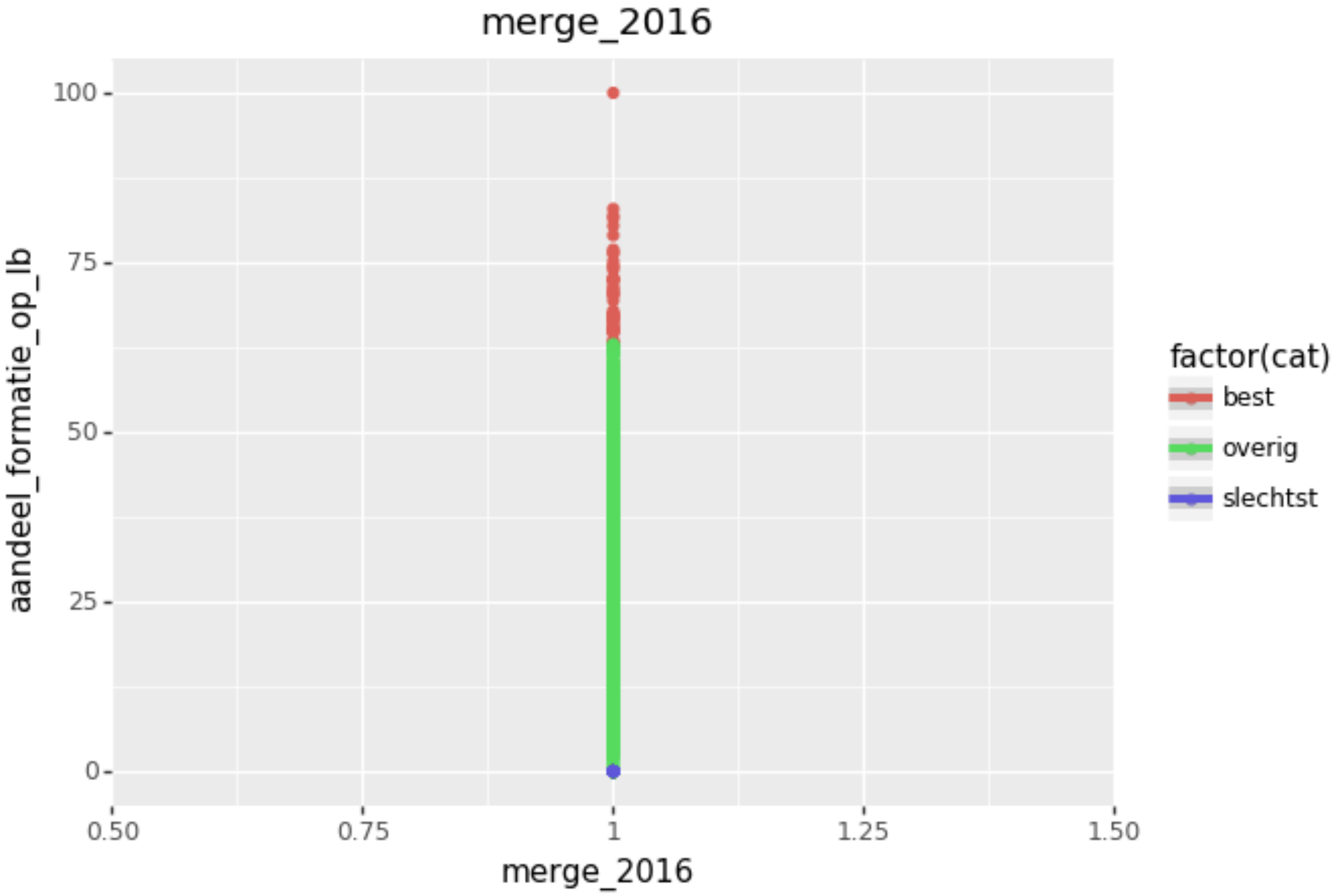


```
<ggplot: (279323996)>
merge_2012
NO NON-NA ROWS
merge_2013
NO NON-NA ROWS
merge_2014
NO NON-NA ROWS
merge_2015
NO NON-NA ROWS
OWBGNUMMER
```

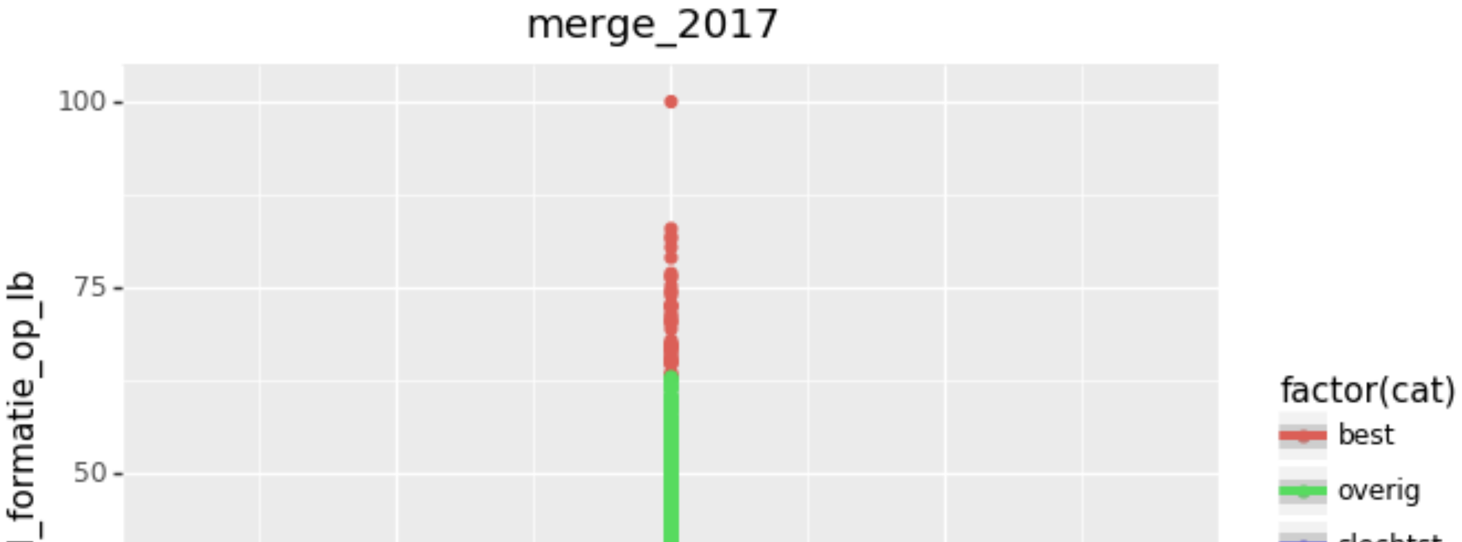


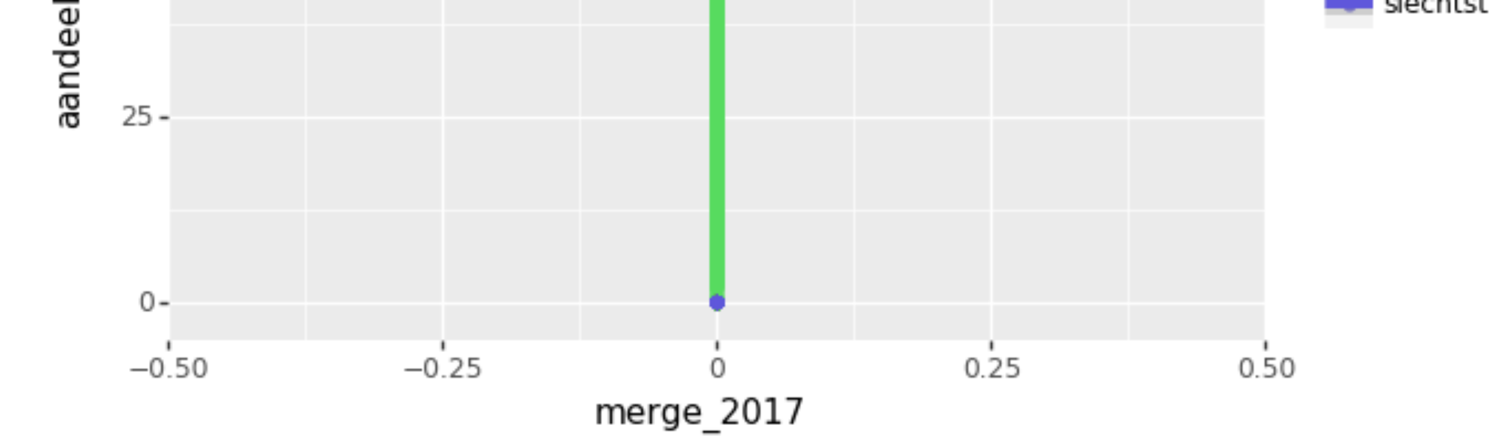


```
<ggplot: (286755094)>
merge_2016
```

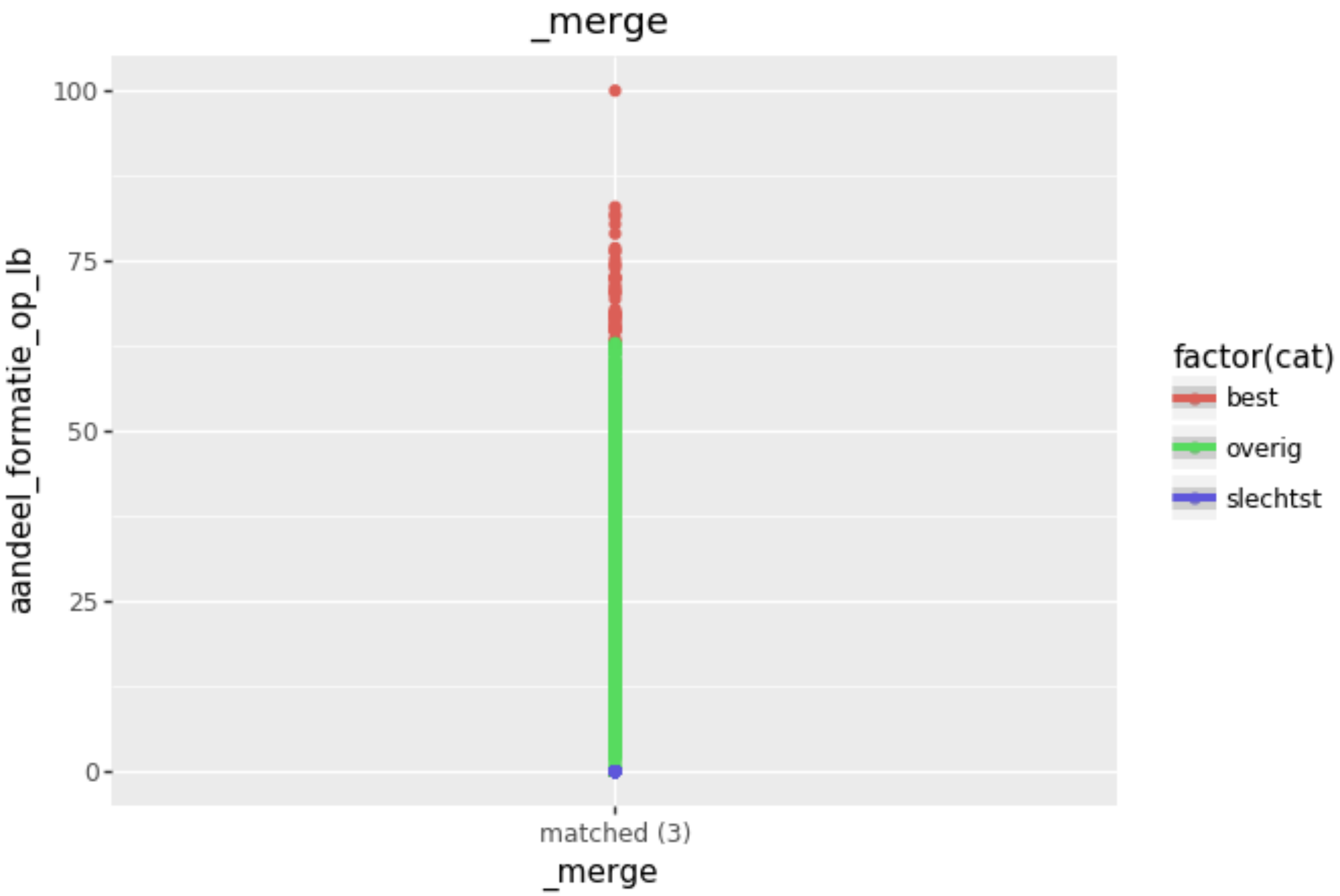


```
<ggplot: (289912959)>
merge_2017
```

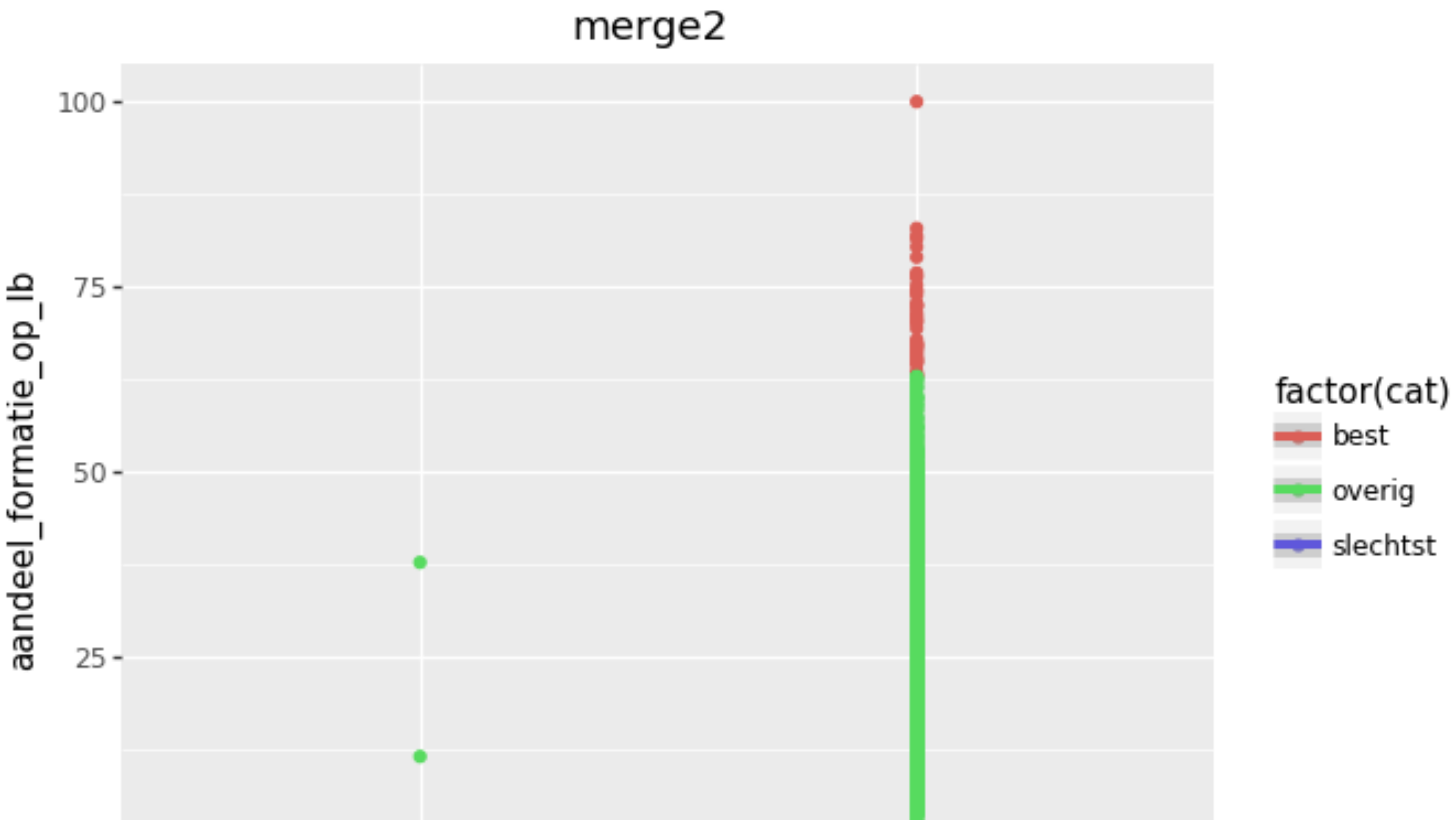




```
<ggplot: (290514615)>
_merge
```



```
<ggplot: (-9223372036564261253)>
merge2
```



```
<ggplot: (285906538)>
```

Out[1]:

```
[]
```

In [ ]: