1. **Stacked bar chart**

+ Clearly shows the progression in prediction classifications

+ Shows explicit percentage correctly specified in each interval (can also be removed if you don’t like it)

- Doesn’t take into account size of each interval

Chart, bar chart

Description automatically generated

1. **Bar chart**

+ Clearly shows the progression in prediction classifications

- Doesn’t take into account size of each interval

- Less compact than figure (1), but does make it easier to see actual percentages

1. **Bar chart (2)**

+ Same as above, now includes a line for size

1. **Line chart**

+ Maybe shows the progression over all prediction classifications even better intuitively than previous figures, also shows actual percentages even more clearlly

- Doesn’t take into account size of each interval

1. **Line chart (2)**

+ Same as above, now includes size as shaded area

1. **Bubble plot**

+ Laat percentage correctly predicted zien per interval (kan met expliciet getal) + een heel visuele representatie van size

- Laat niet de andere percentages expliciet zien, dus je ziet alleen hier hoeveel goed (en hoeveel fout kan je dan direct uitrekenen), maar niet hoeveel er onder of hoeveel er boven

- De x-as laat nog niet goed zien dat het om intervals gaat, maar als we deze figuur er in willen dan vind ik wel een manier met fotoshop ofzo ☺

1. **Scatter with 45 degree line**

Ik hoop dat dit een beetje overeenkomt met wat jullie in gedachte hadden voor de ‘AUC’-achtige figuur. Ik heb ze gemaakt als volgt. Ik heb eerst 160 intervals gemaakt in stapjes van 2, dus [0,2), [2,4), …, [316,318), op basis van de werkelijke waarde. Vervolgens heb ik de mediaan/gemiddelde uitgerekend van de werkelijke waarde in dit interval (y-as) en dit uitgezet tegen de mediaan/gemiddeldes van de voorspelde waarden voor deze observaties.

* Allemaal hetzelfde patroon
* Allemaal ondervoorspellen
* XGBoost beste model?
* Gat tussen 25 – 50
* Hurdle model / zero-inflated model 🡪 further research
* Tweedie meeste non-linearity, maar lijkt niet per se beste voor waardes > 0. Meeste spreiding. Minste spreiding MeXGBoost.
* [75,125] interval highest density als niet nul 🡪 maken van histogram + density plot
* ME boeiend bij xgboot, minder bij ann

* **Grotere stappen (10,20)**
* **Proberen te transponeren**
* Zwarte letters
* Kijken gridlijnen
* **Percentielen (laagste 1% van je echte obversaties [x-as 1-100], …) sorteren echte data, en daarbinnen sorteren op predicted variable**
* **Lijntjes erbij met min & max of interquartile range (bij grotere buckets)**

Ik heb ze ook gemaakt gebaseerd op een interval van 4 mug/g, waardoor er minder punten in zitten, en het misschien nog net weer wat duidelijker wordt. De general pattern is hetzelfde in deze figuren als hierboven. Ik heb hier ook de x-as aangepast zodat er ook lijnen zitten bij 25, 75, 125, etc.

1. **Updated heatmaps**

*ANN*

**

*MeANN*

**

*XGBoost*

**

*MeXGBoost*

**

*TweedieXGBoost*



1. **Heatmap transpose**

*ANN*

******