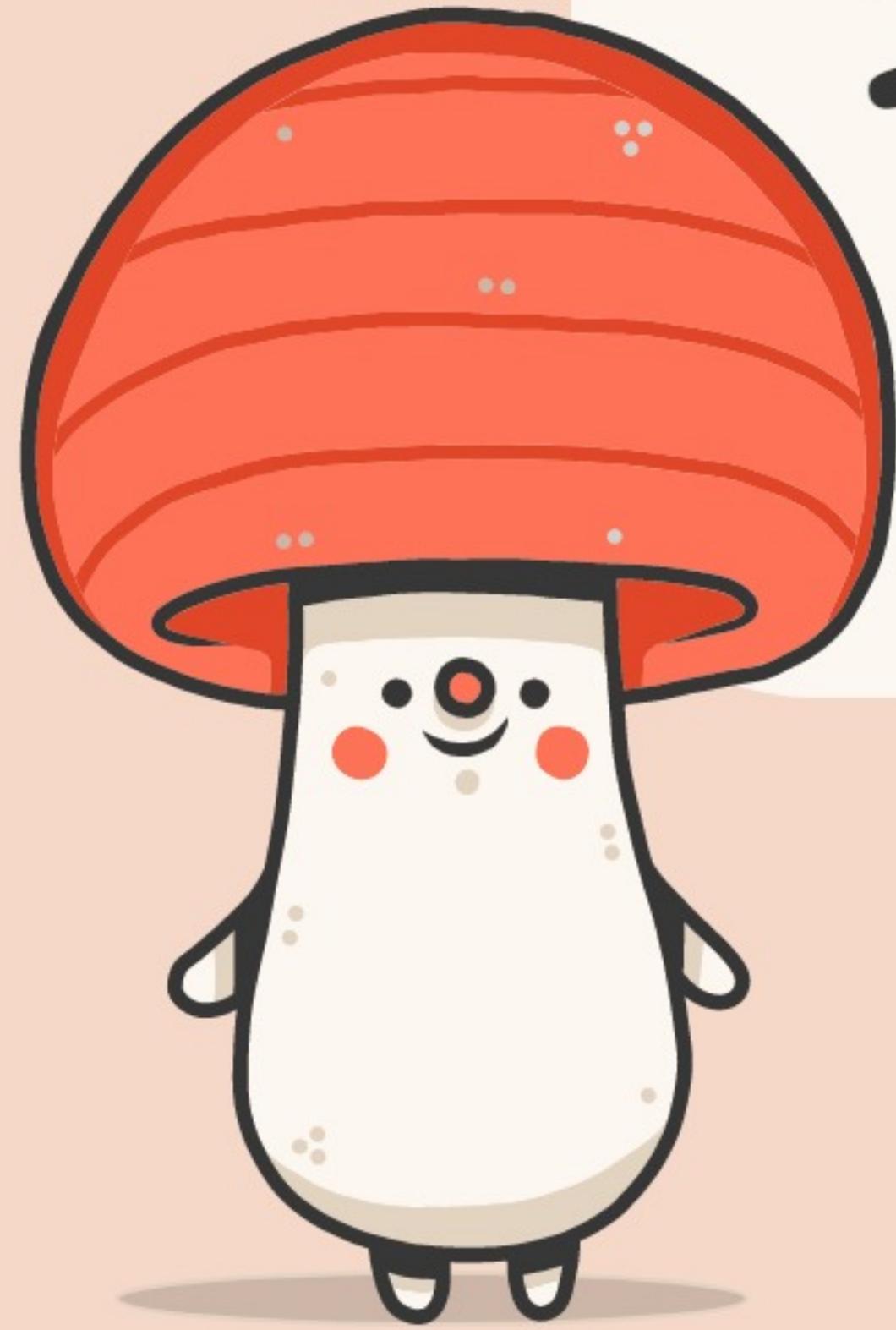


分類菇可不可食用



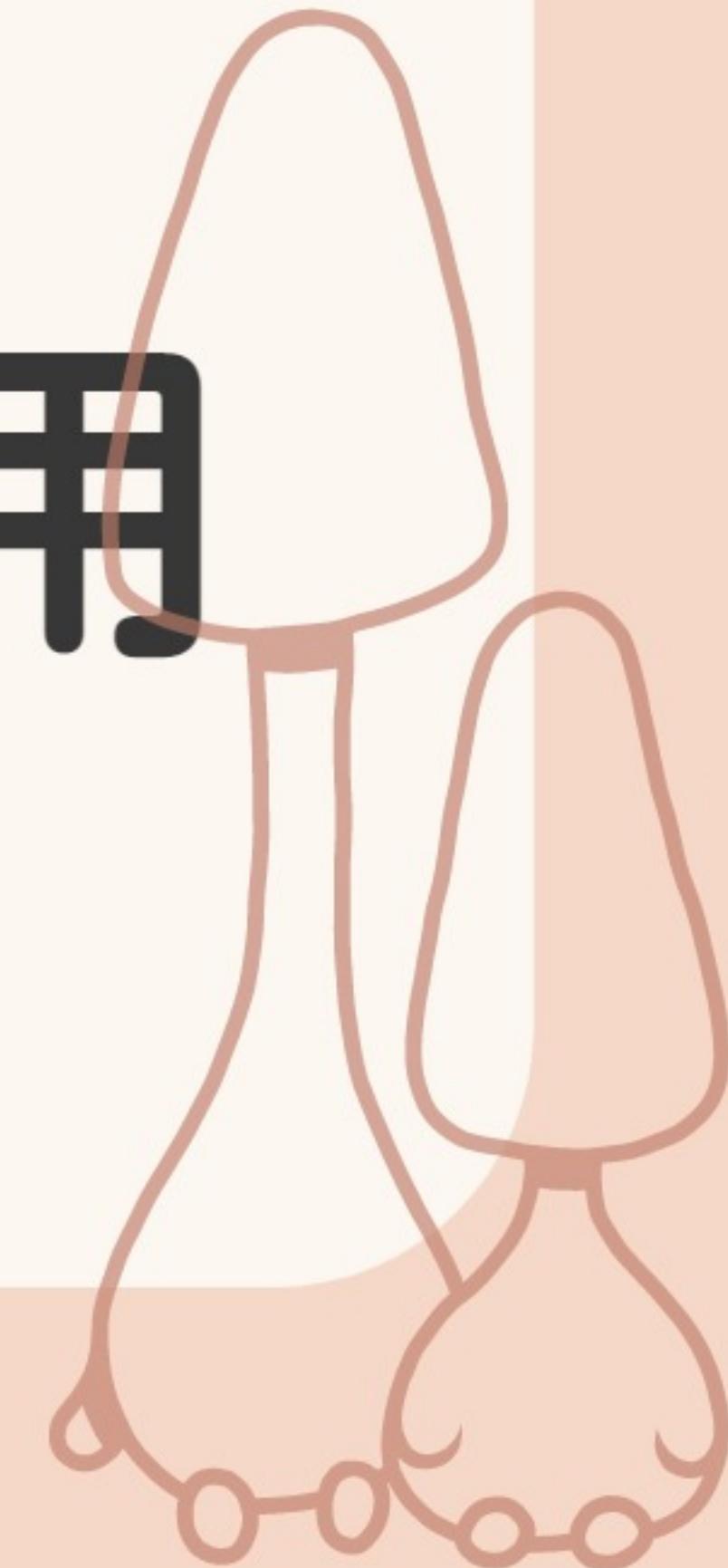
邱婉如



邱繼賢



廖偉傑



Data Overview & Goal



研究目的：

分辨蘑菇是否可食用 (class : **poisonous=p, edible=e**)



記錄了 **61069** 株蘑菇的外觀及生長環境特徵，總共 **21** 個變數。

Data Pre-processing



移除變數 (>40%)



移除觀測值 (<10%)



補值

剩餘 15 個變數 58598 筆觀測值

變數名稱	遺失比例
veil.type	94.8%
spore.print.color	89.6%
veil.color	87.9%
stem.root	84.4%
stem.surface	64.4%
gill.spacing	41.0%
cap.surface	23.1%
gill.attachment	16.2%
ring.type	4.0%

Missing Value Imputation



class	cap.surface	gill.attachment	其他 12 個變數
e	t	NA	
:	:	:	
p	g	f	



p	NA	e	
p	NA	NA	
e	NA	NA	

Missing Value Imputation



class	cap.surface	gill.attachment	其他 12 個變數
e	t	NA	
:	:	:	
p	g	f	

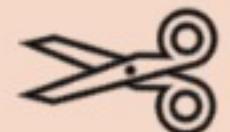


p	w	e	
p	g	NA	
e	h	NA	

Missing Value Imputation



class	cap.surface	gill.attachment	其他 12 個變數
p	w	e	
:	:	:	
p	g	f	

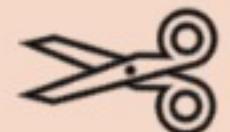


e	t	NA	
p	g	NA	
e	h	NA	

Missing Value Imputation

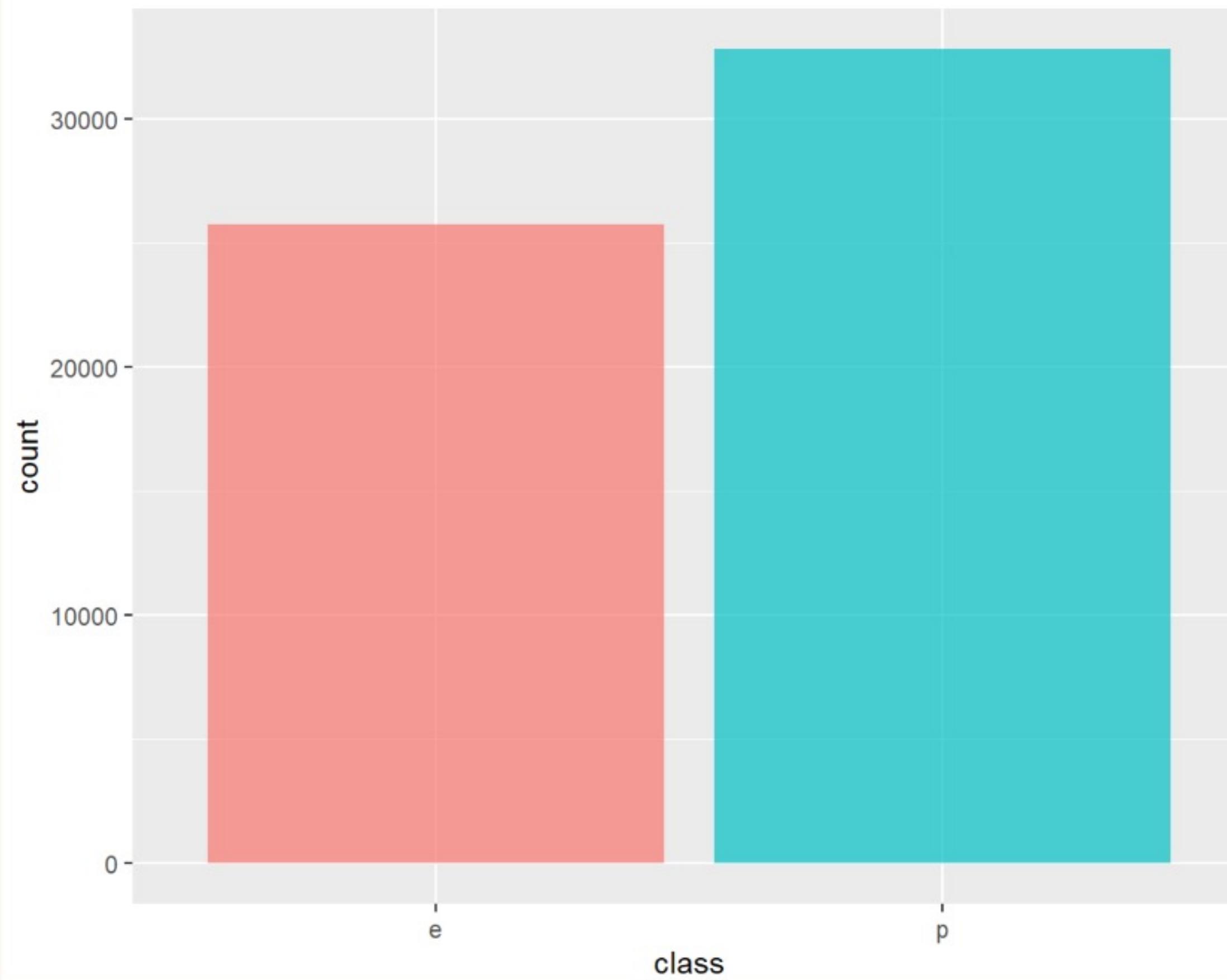


class	cap.surface	gill.attachment	其他 12 個變數
p	w	e	
:	:	:	
p	g	f	



e	t	a	
p	g	x	
e	h	d	

EDA : Response



有毒蘑菇佔 56.2%

class
e
p



Predictors



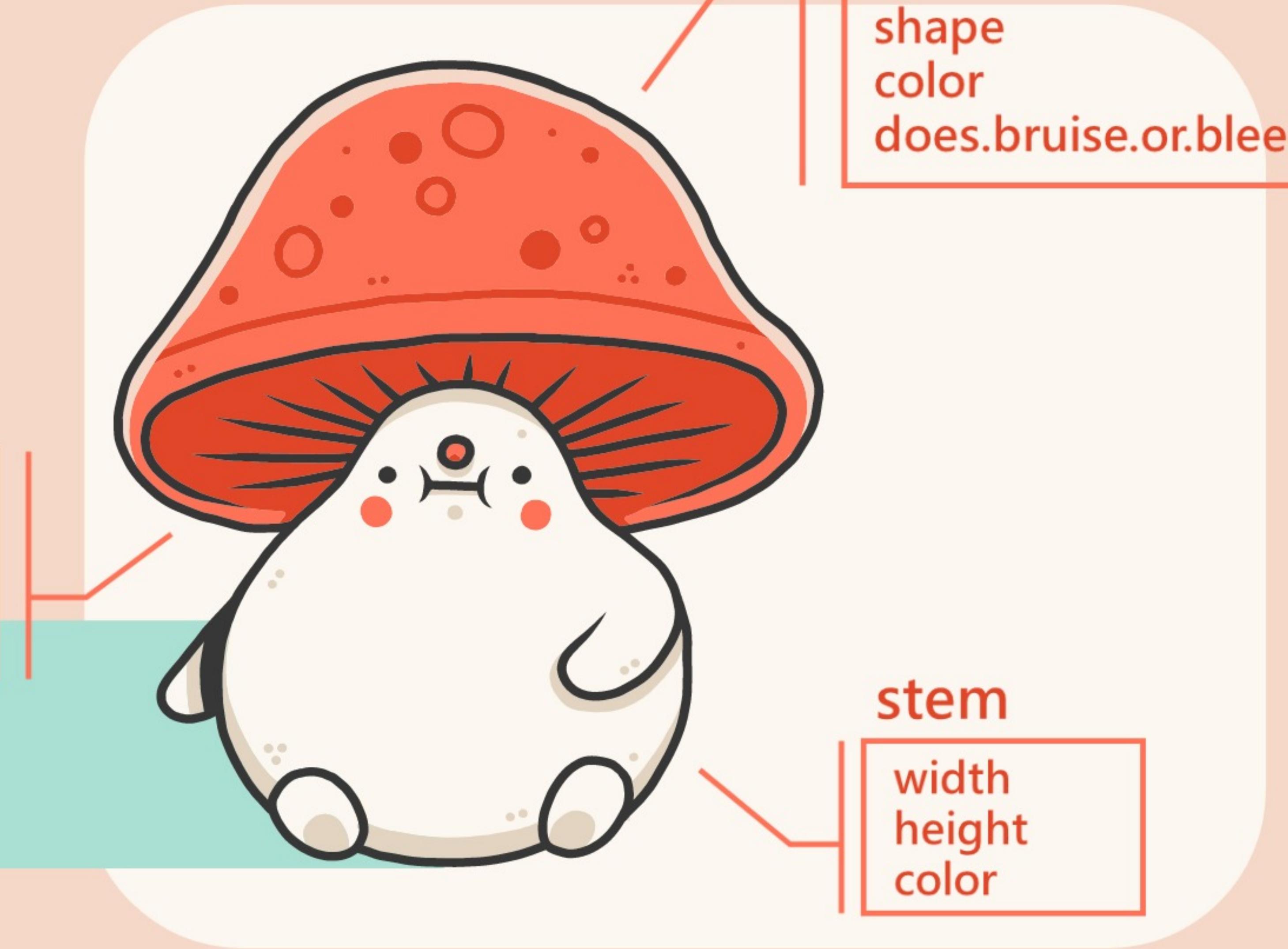
habitat



season

gill

color
attachment
has.ring
ring.type

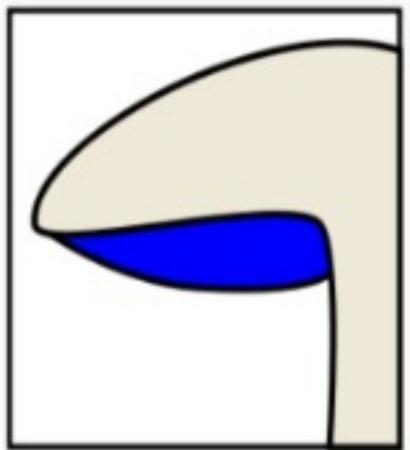


cap

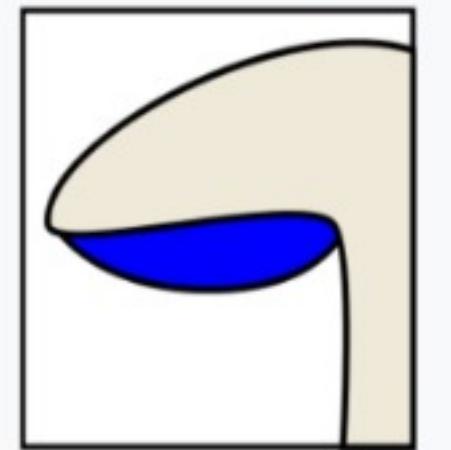
diameter
surface
shape
color
does.bruise.or.bleed

stem

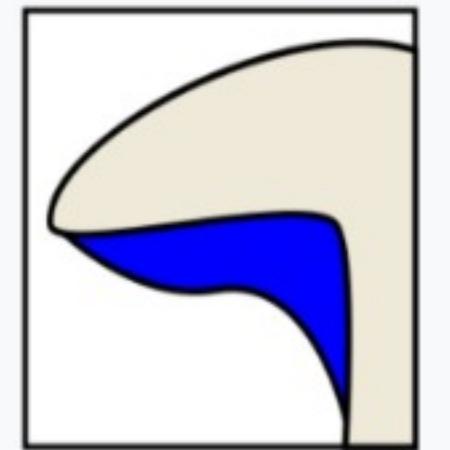
width
height
color



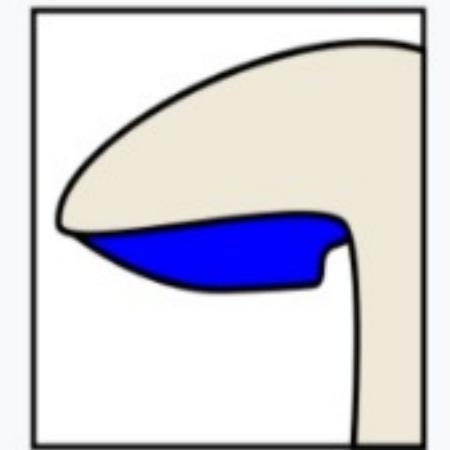
Adnate



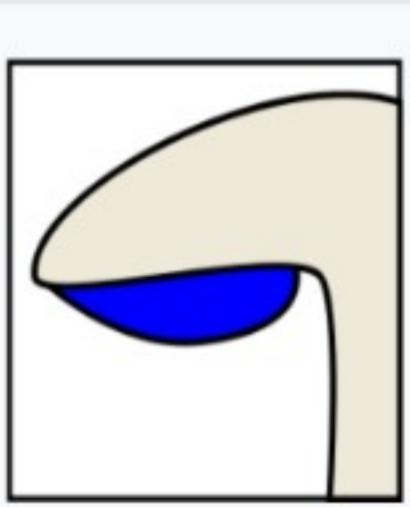
Adnexed



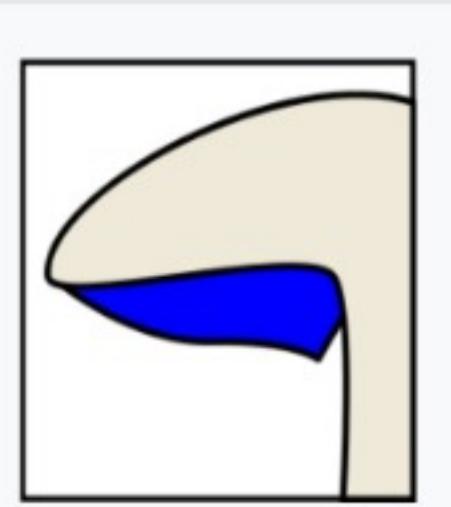
Decurrent



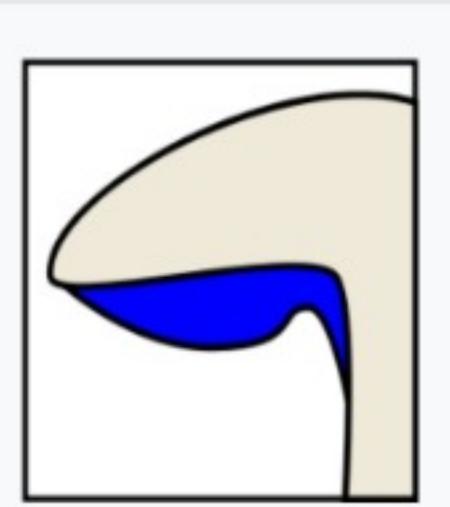
Emarginate



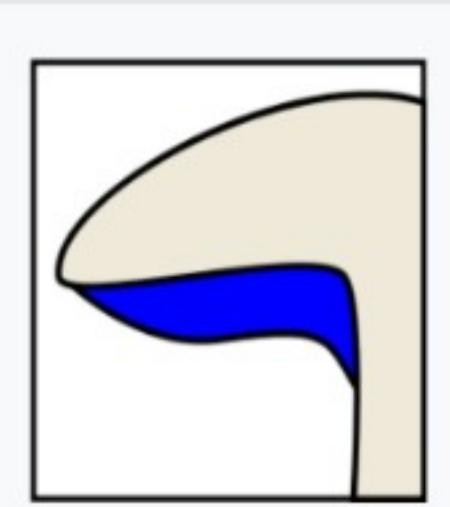
Free



Seceding



Sinuate



Subdecurrent



cap

diameter
surface
shape
color
does.bruise.or.bleed

gill

color
attachment
has.ring
ring.type



habitat

stem

width
height
color

Predictors



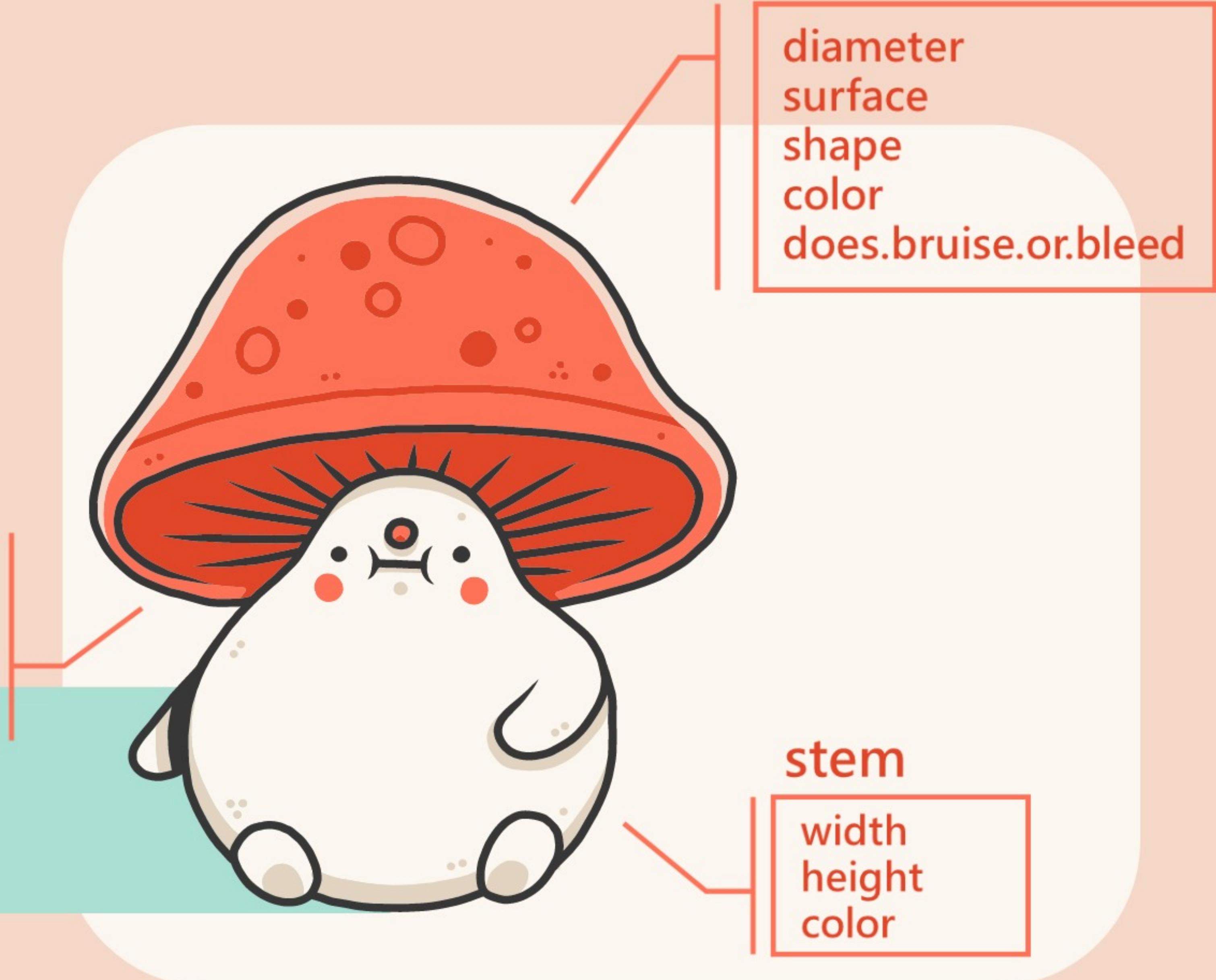
habitat

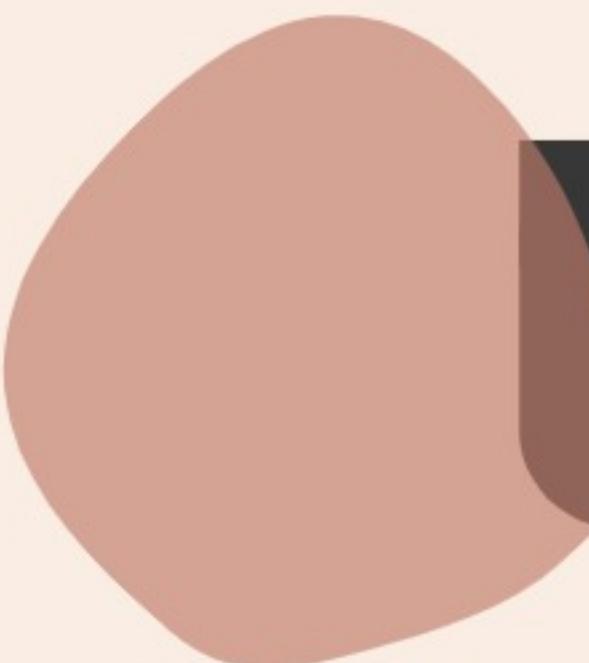


season

gill

color
attachment
has.ring
ring.type





cap.diameter



season



gill

color
attachment
has.ring
ring.type



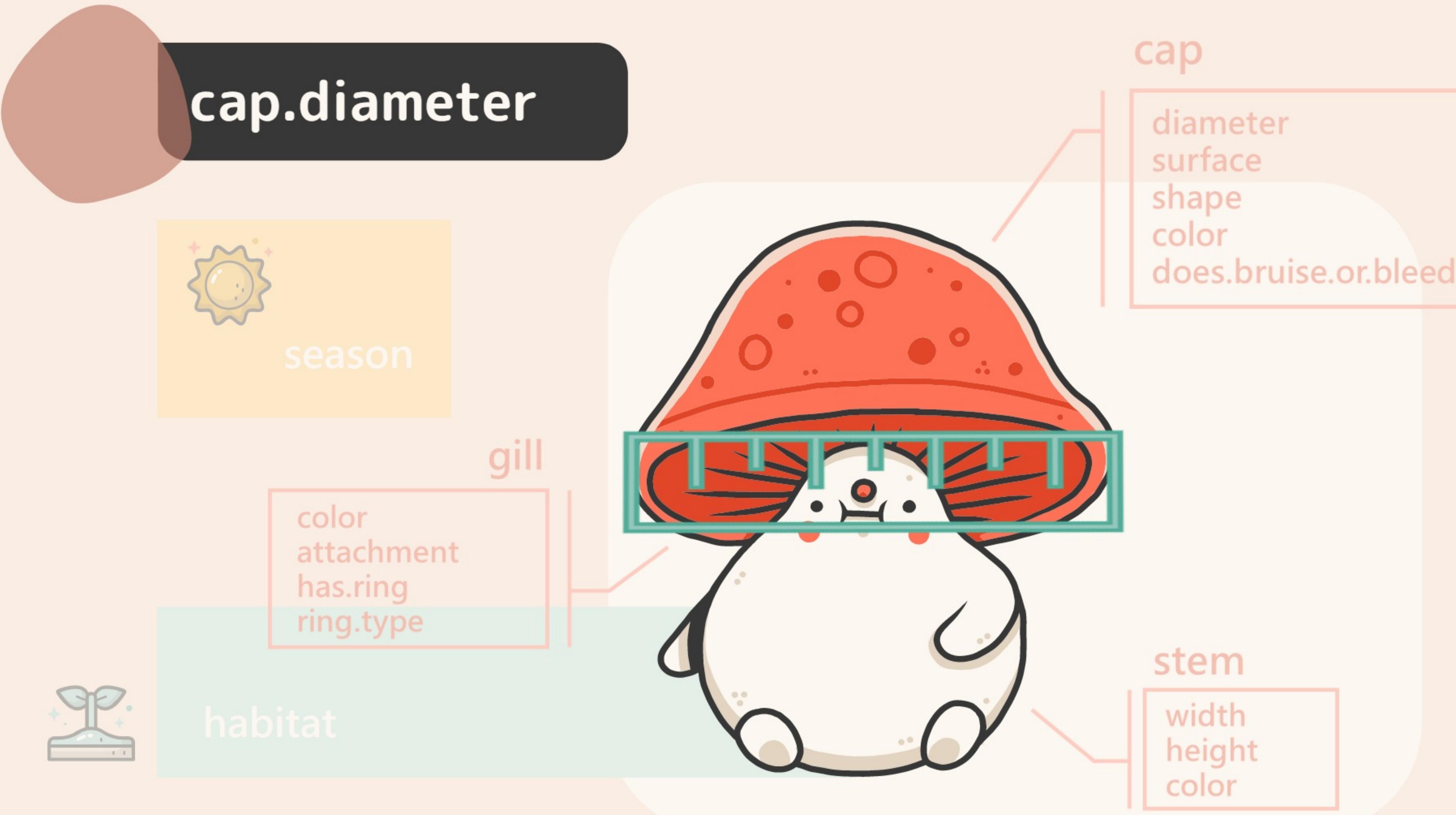
habitat

cap

diameter
surface
shape
color
does.bruise.or.bleed

stem

width
height
color



cap.diameter



season

gill

color
attachment
has.ring
ring.type



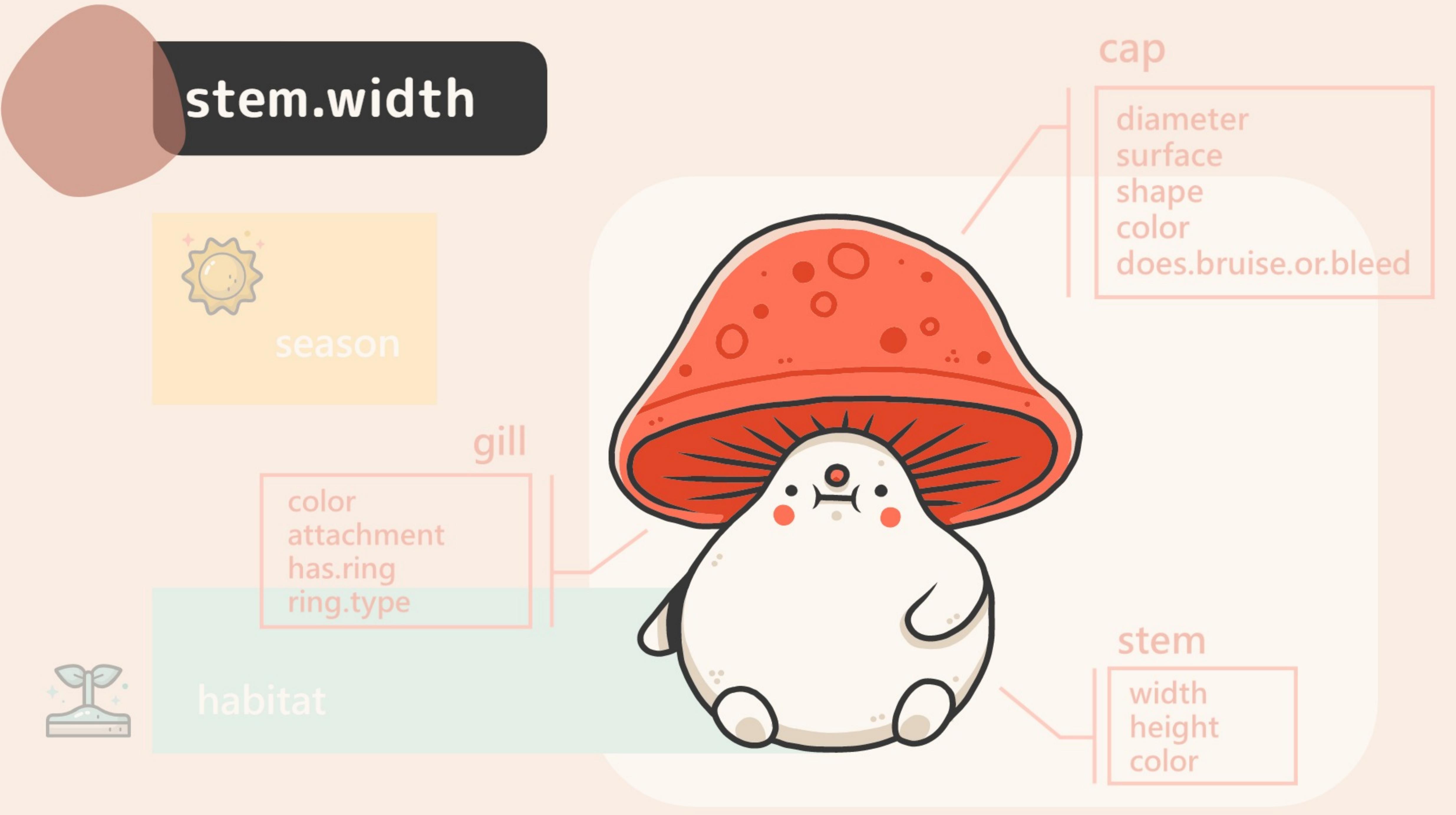
habitat

cap

diameter
surface
shape
color
does.bruise.or.bleed

stem

width
height
color



stem.width



season

gill

color
attachment
has.ring
ring.type



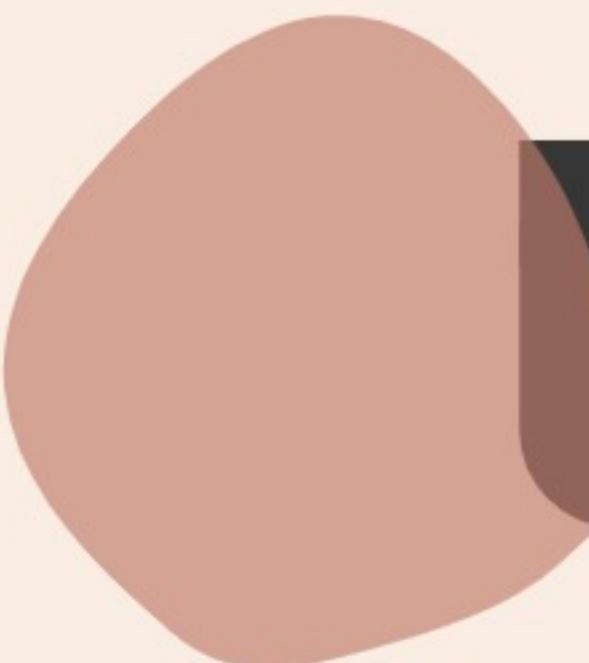
habitat

cap

diameter
surface
shape
color
does.bruise.or.bleed

stem

width
height
color



stem.width



season

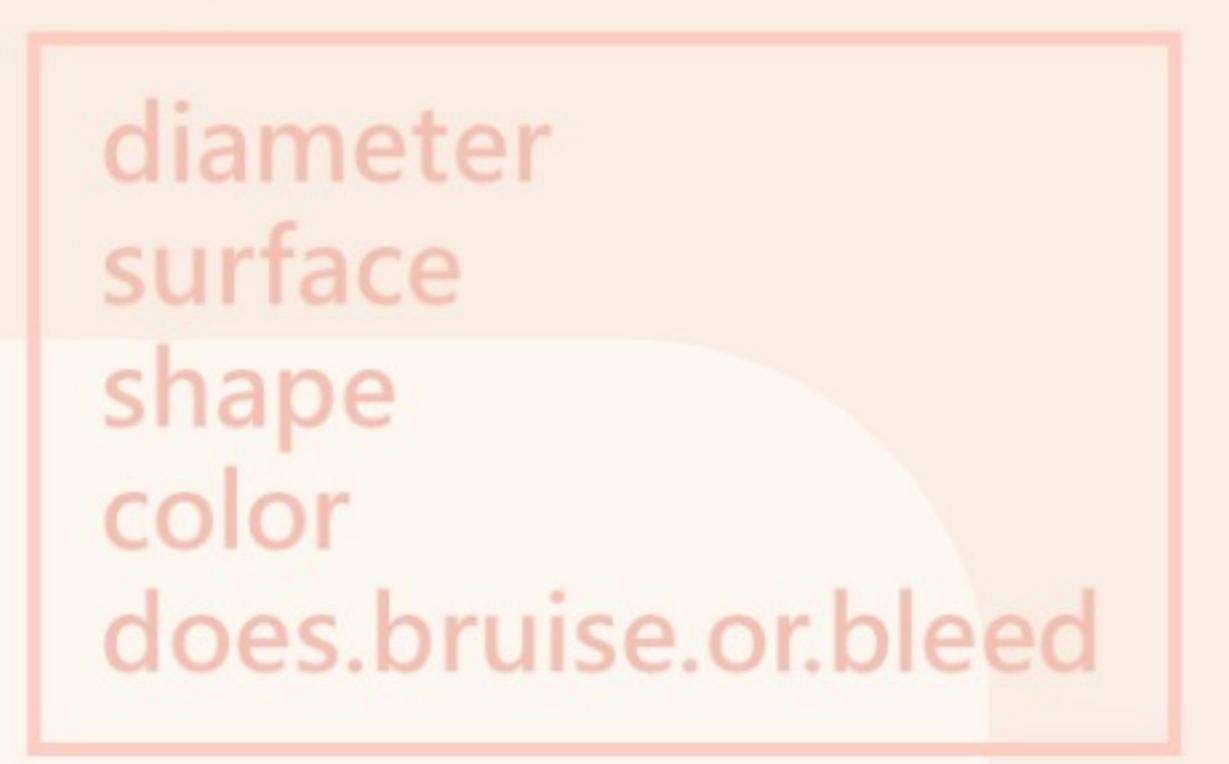


gill

color
attachment
has.ring
ring.type

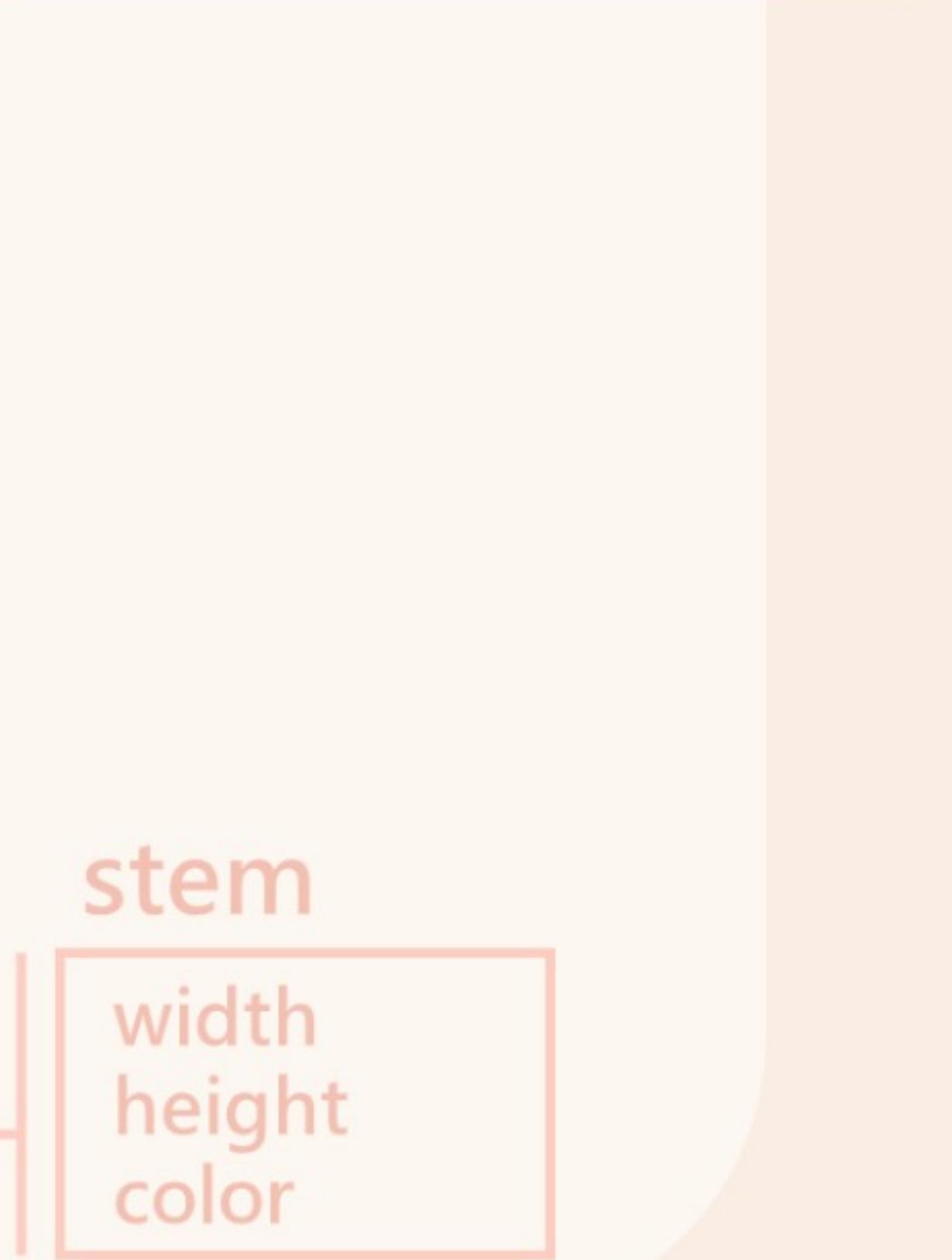


habitat



cap

diameter
surface
shape
color
does.bruise.or.bleed



stem

width
height
color



stem.height



gill

color
attachment
has.ring
ring.type

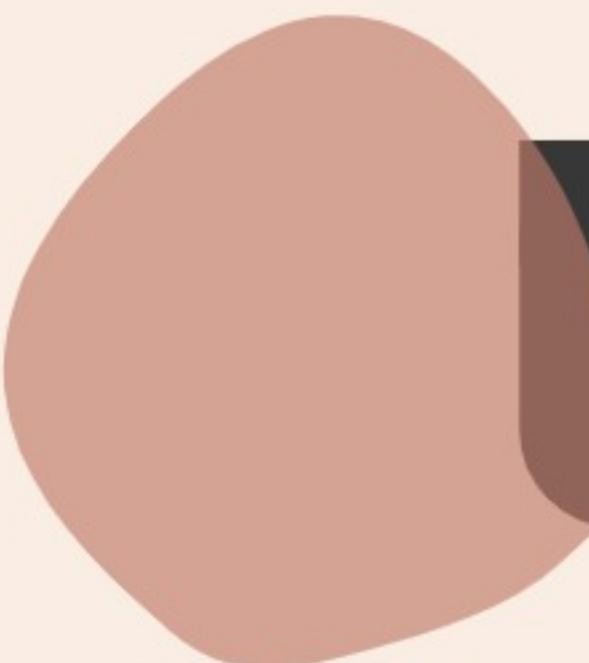


cap

diameter
surface
shape
color
does.bruise.or.bleed

stem

width
height
color



stem.height



season



gill

color
attachment
has.ring
ring.type

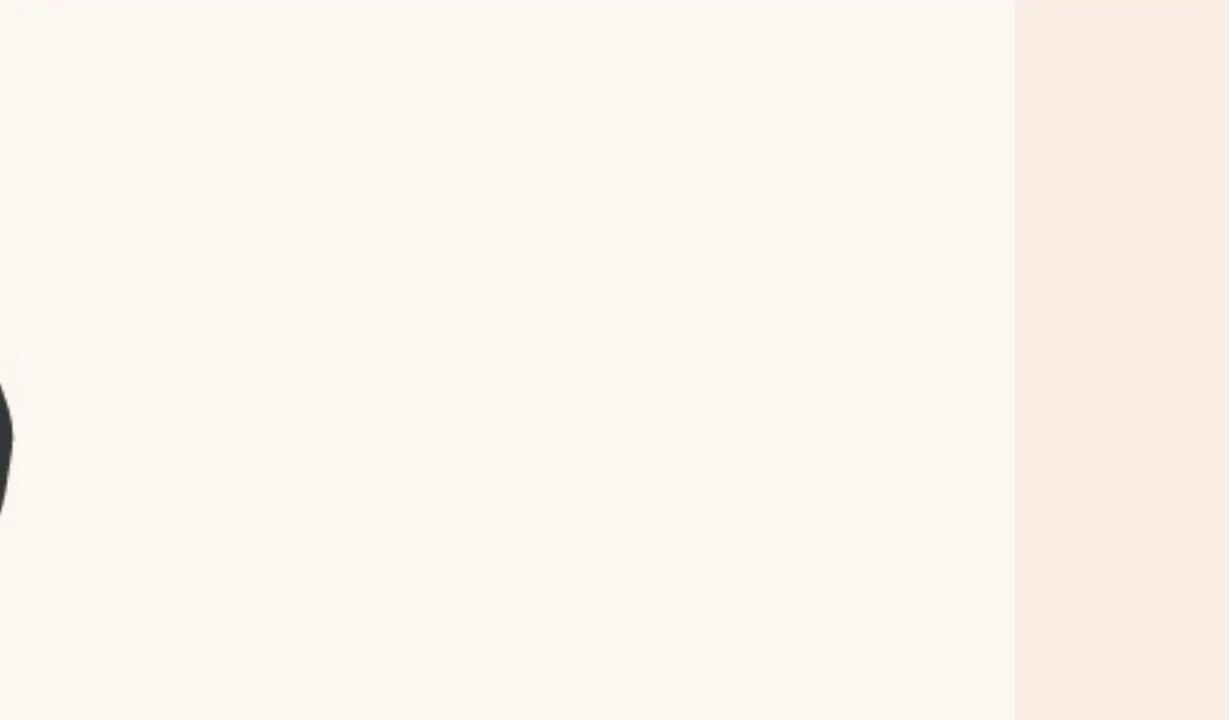


habitat



cap

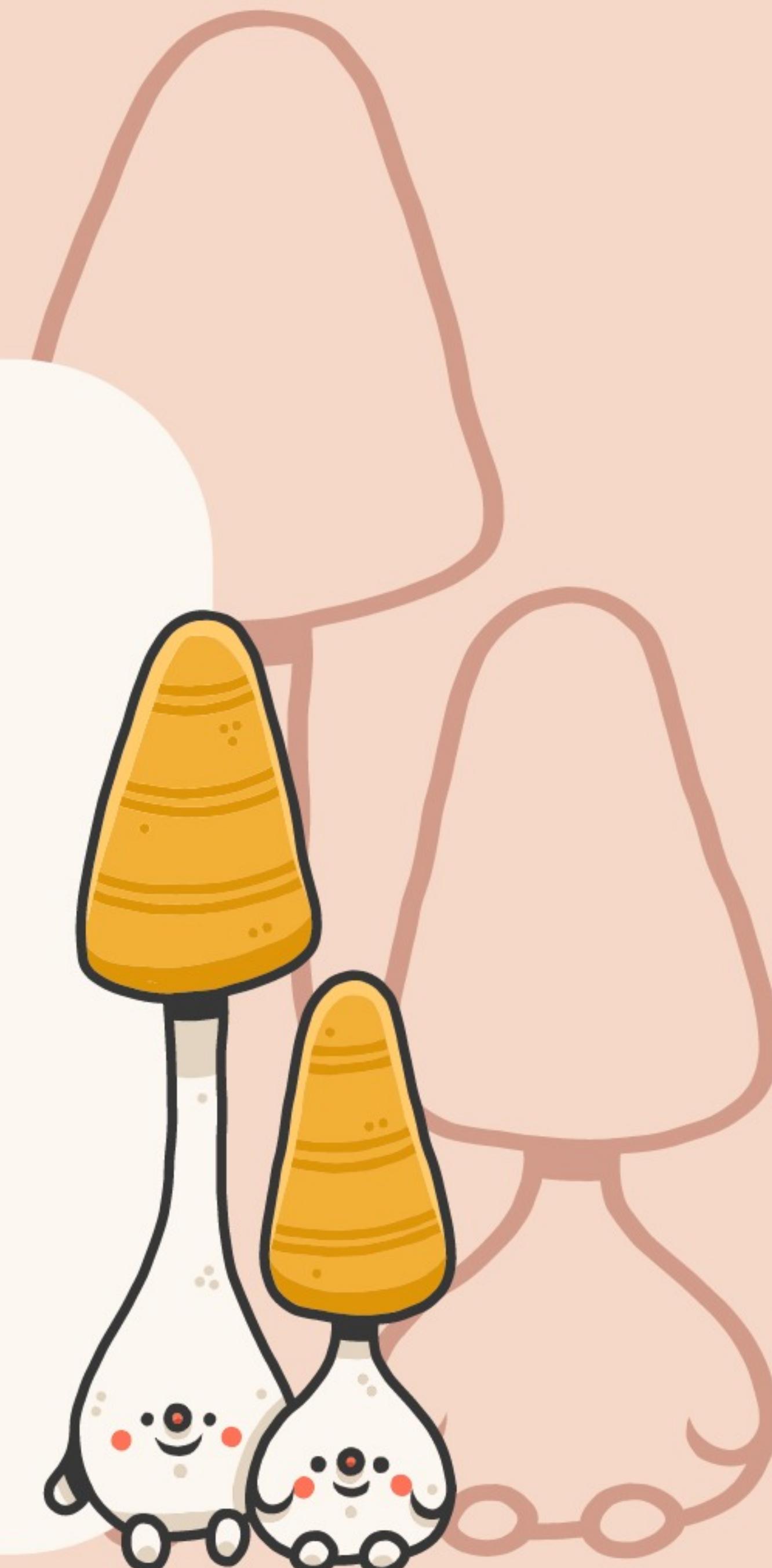
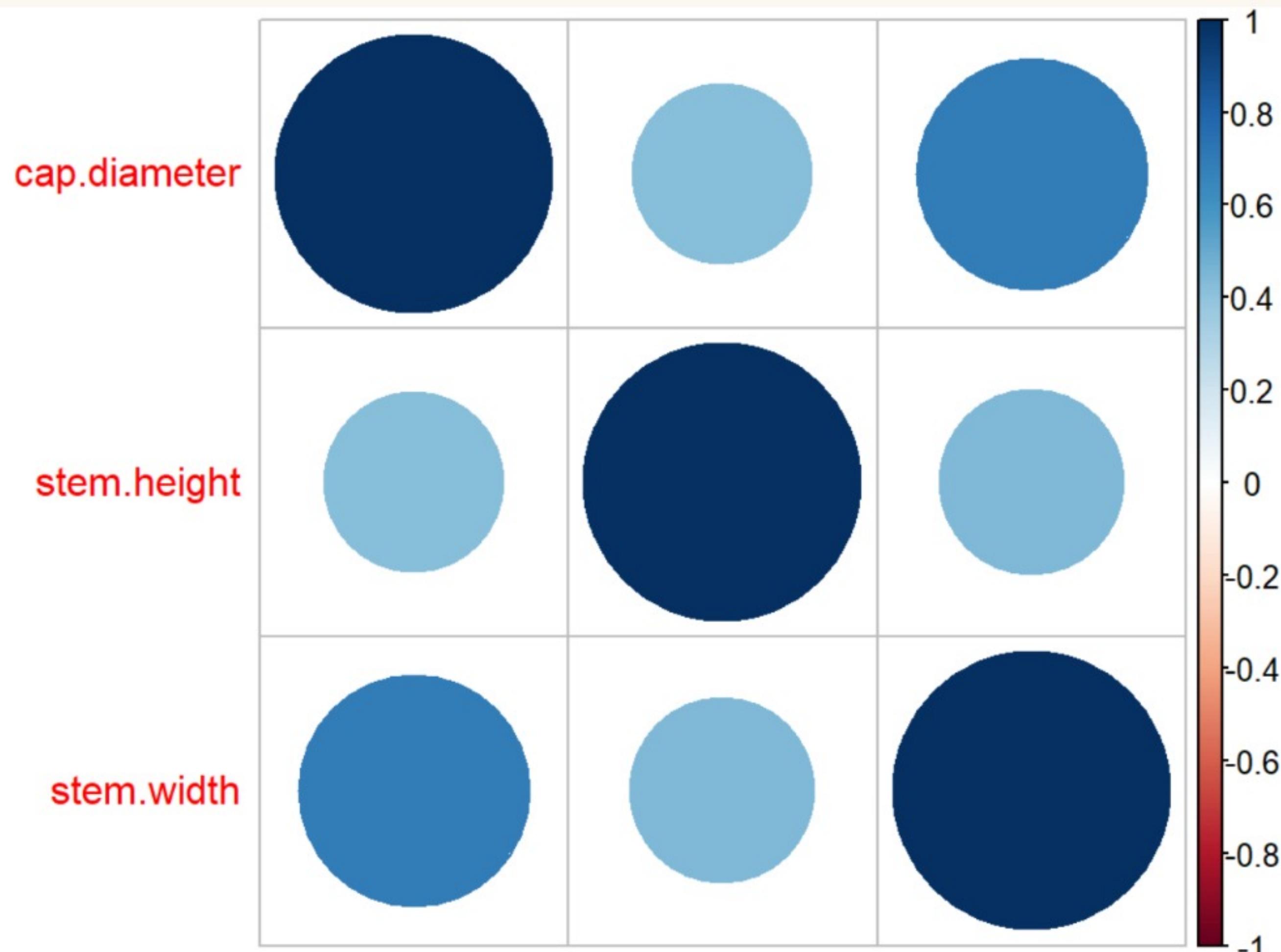
diameter
surface
shape
color
does.bruise.or.bleed



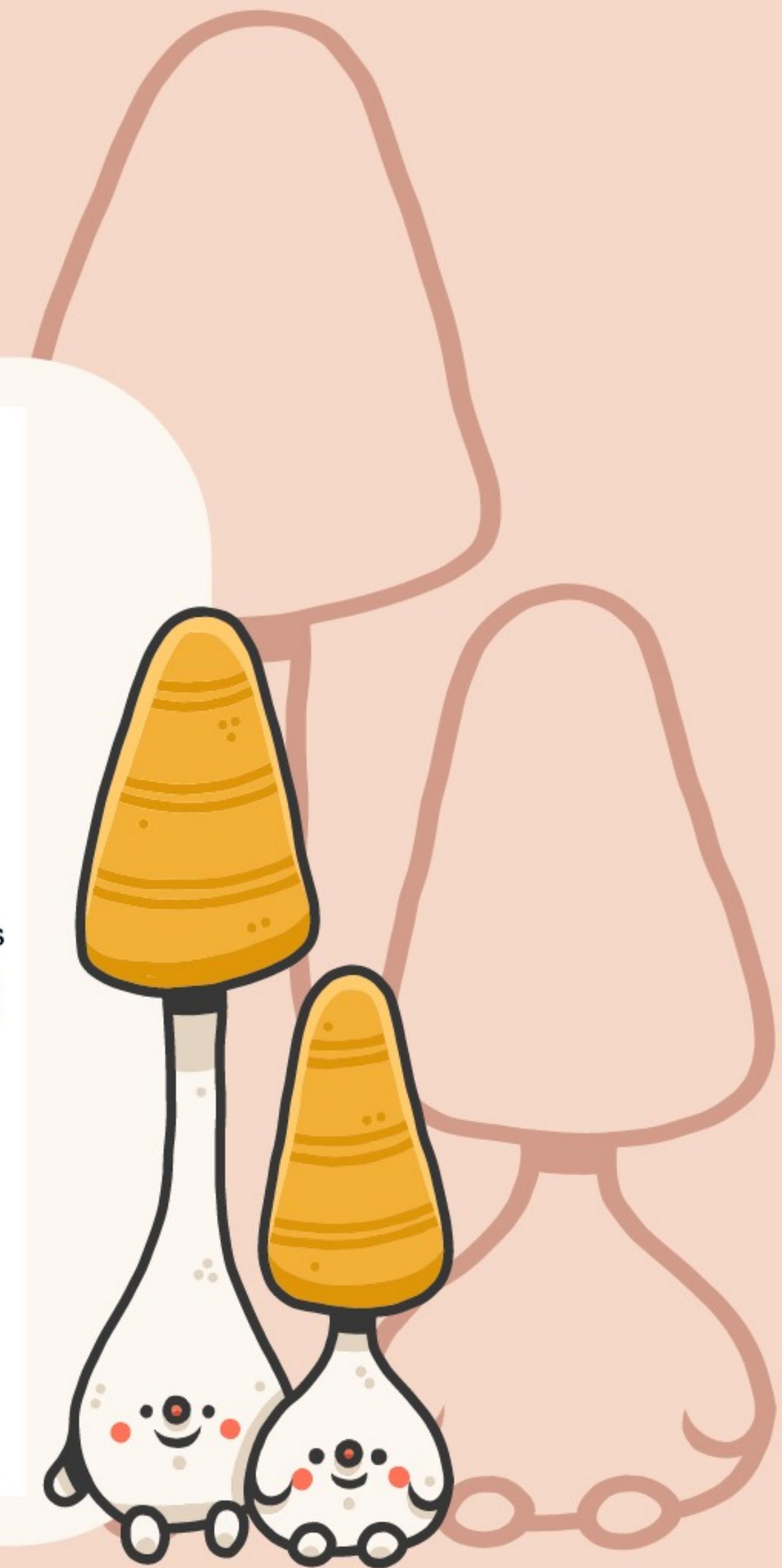
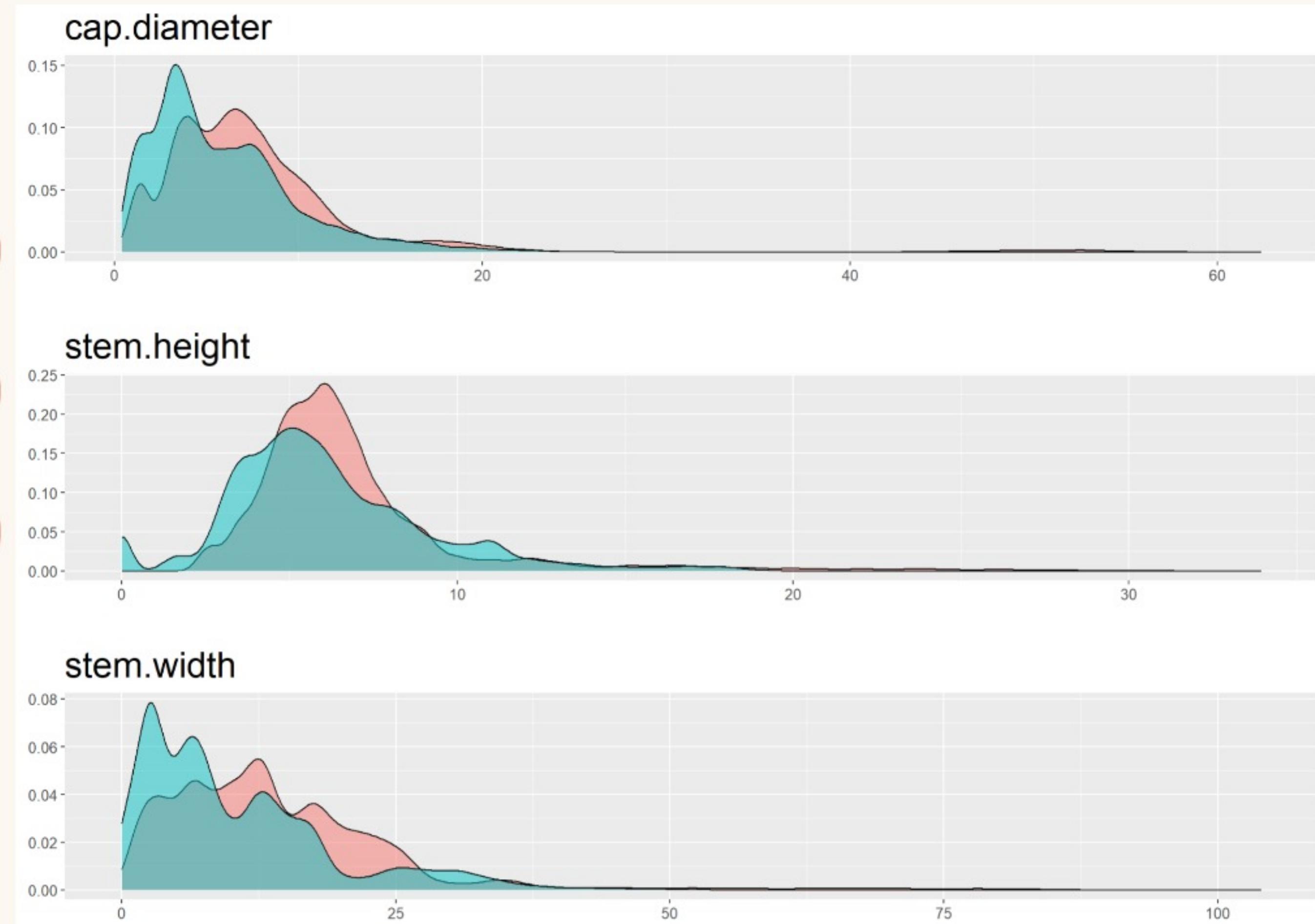
stem

width
height
color

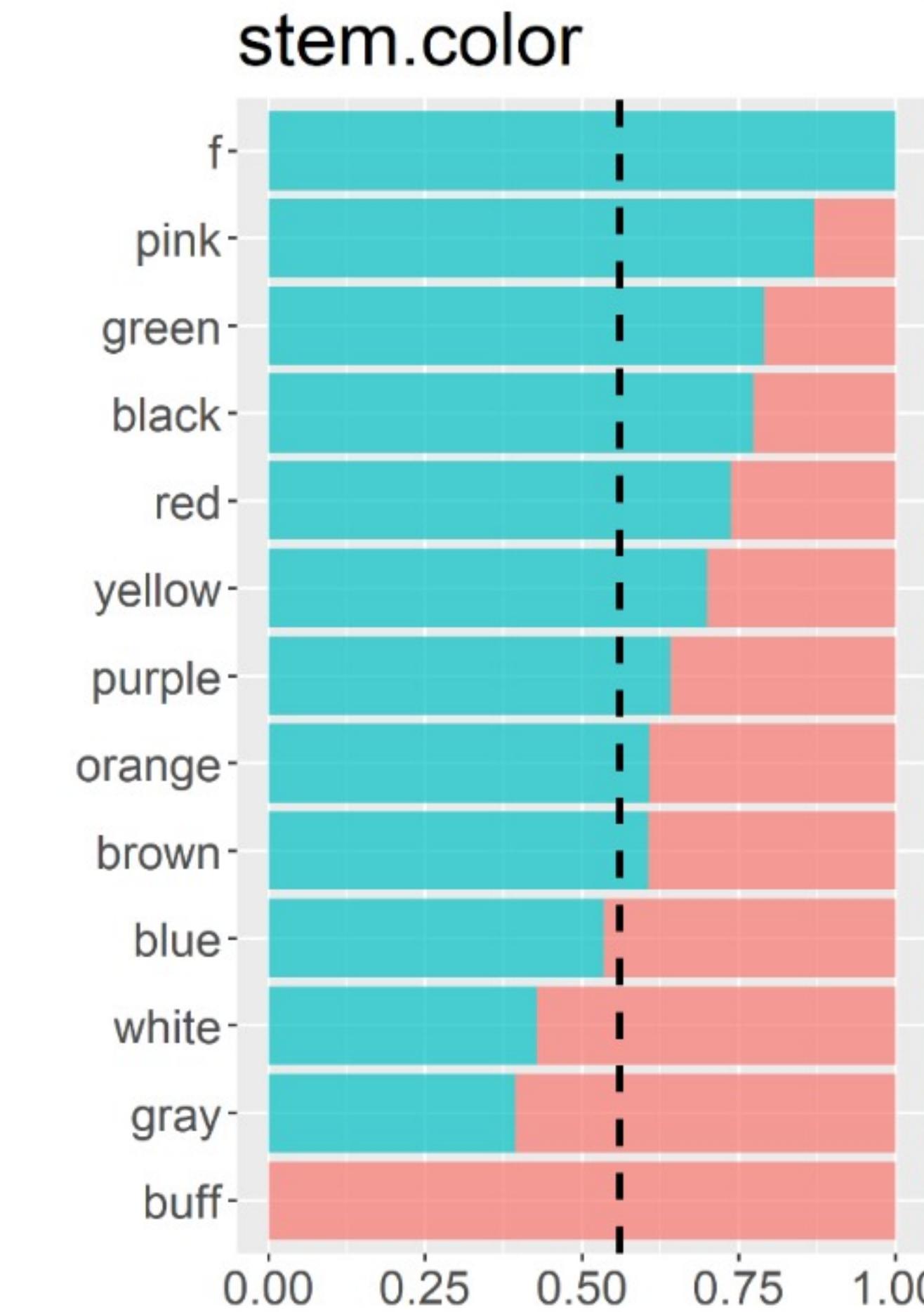
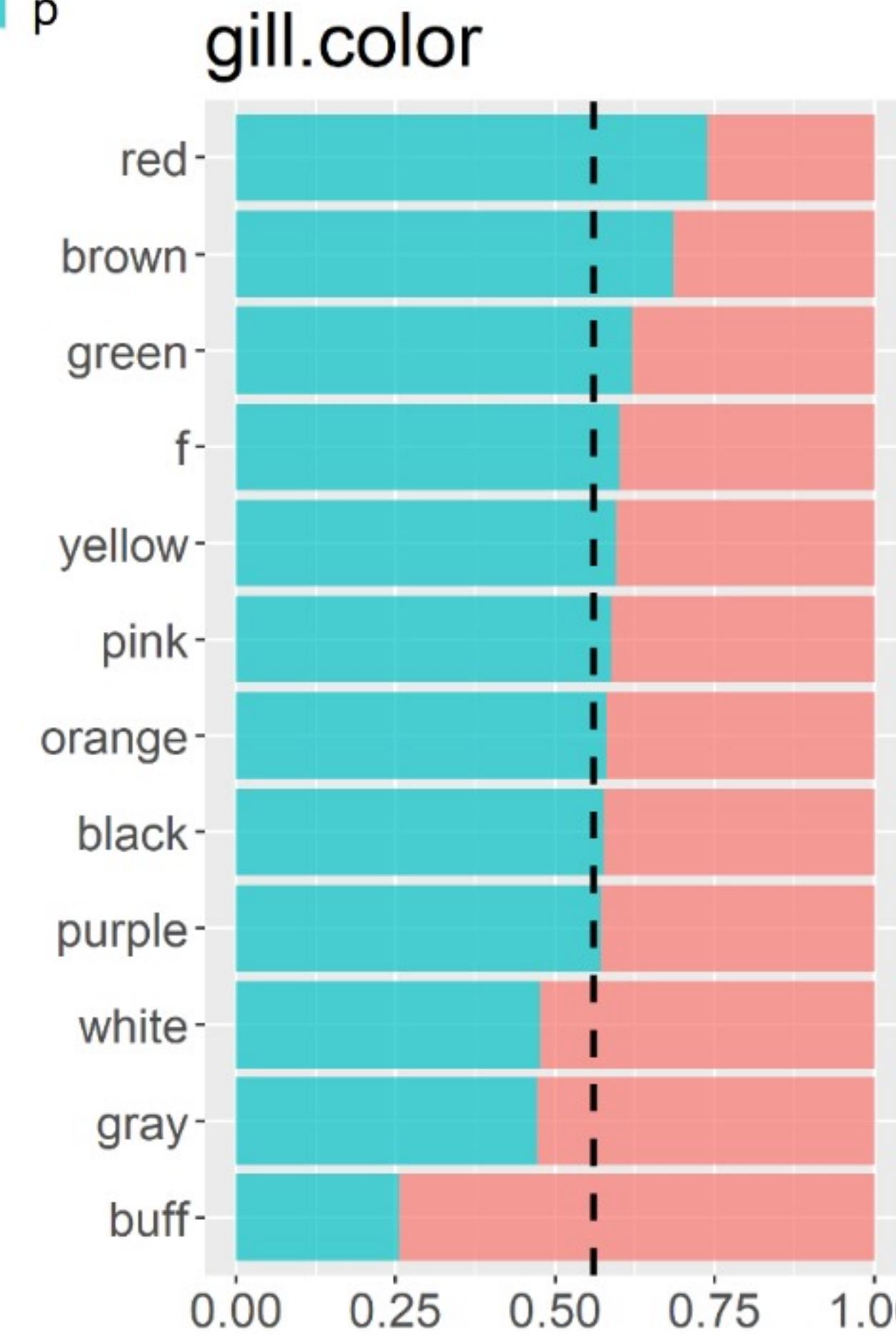
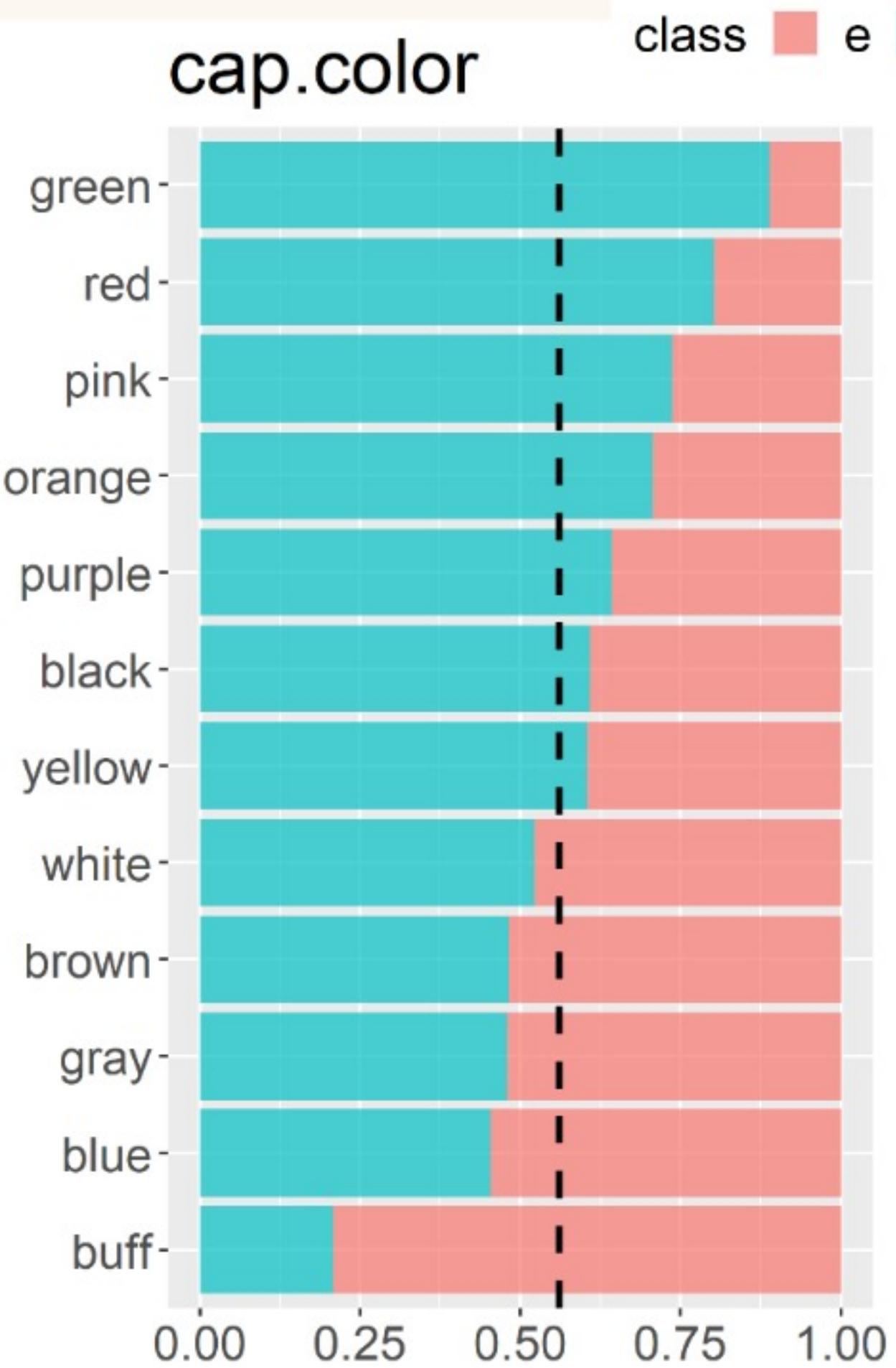
EDA : Size of Mushroom



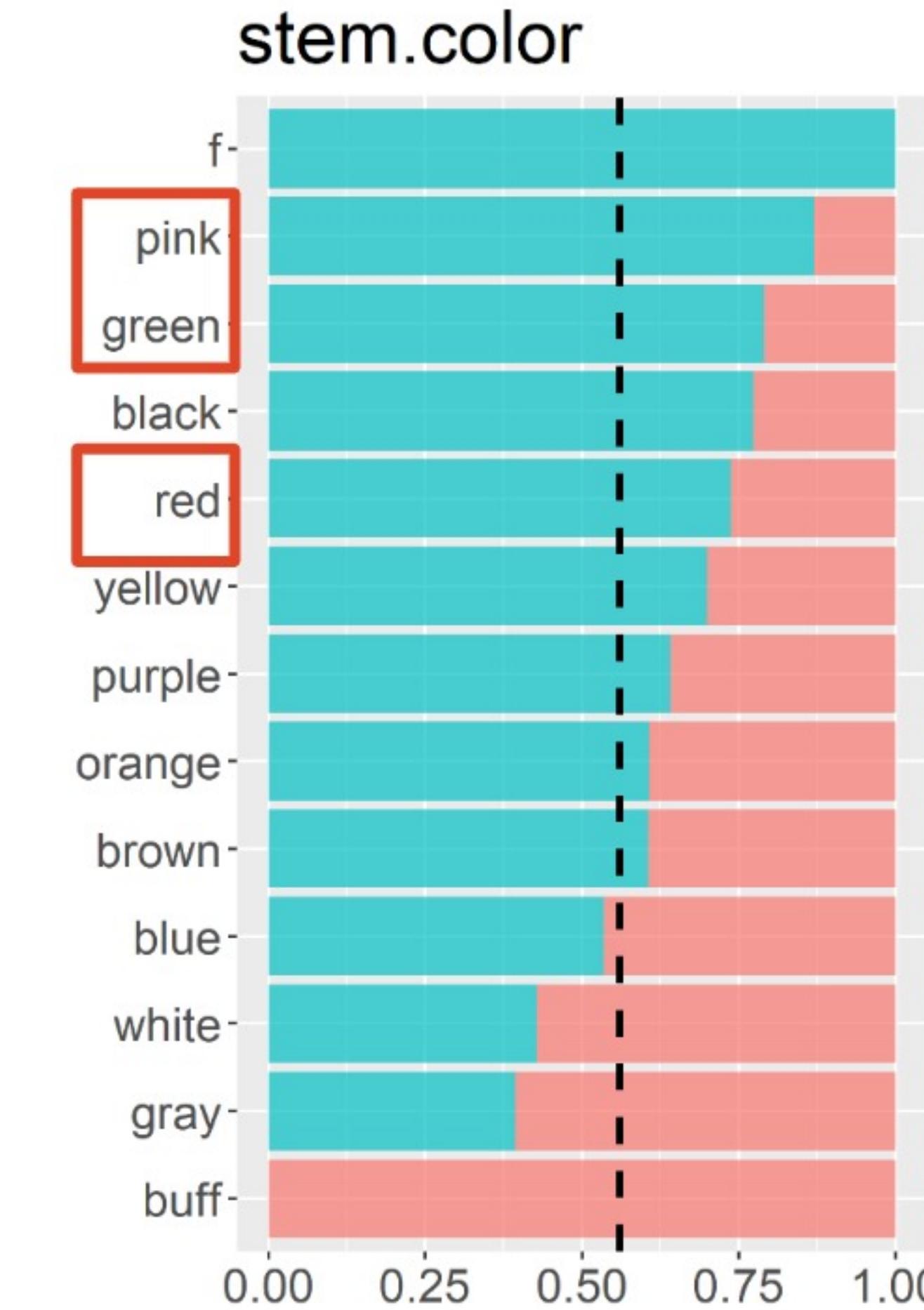
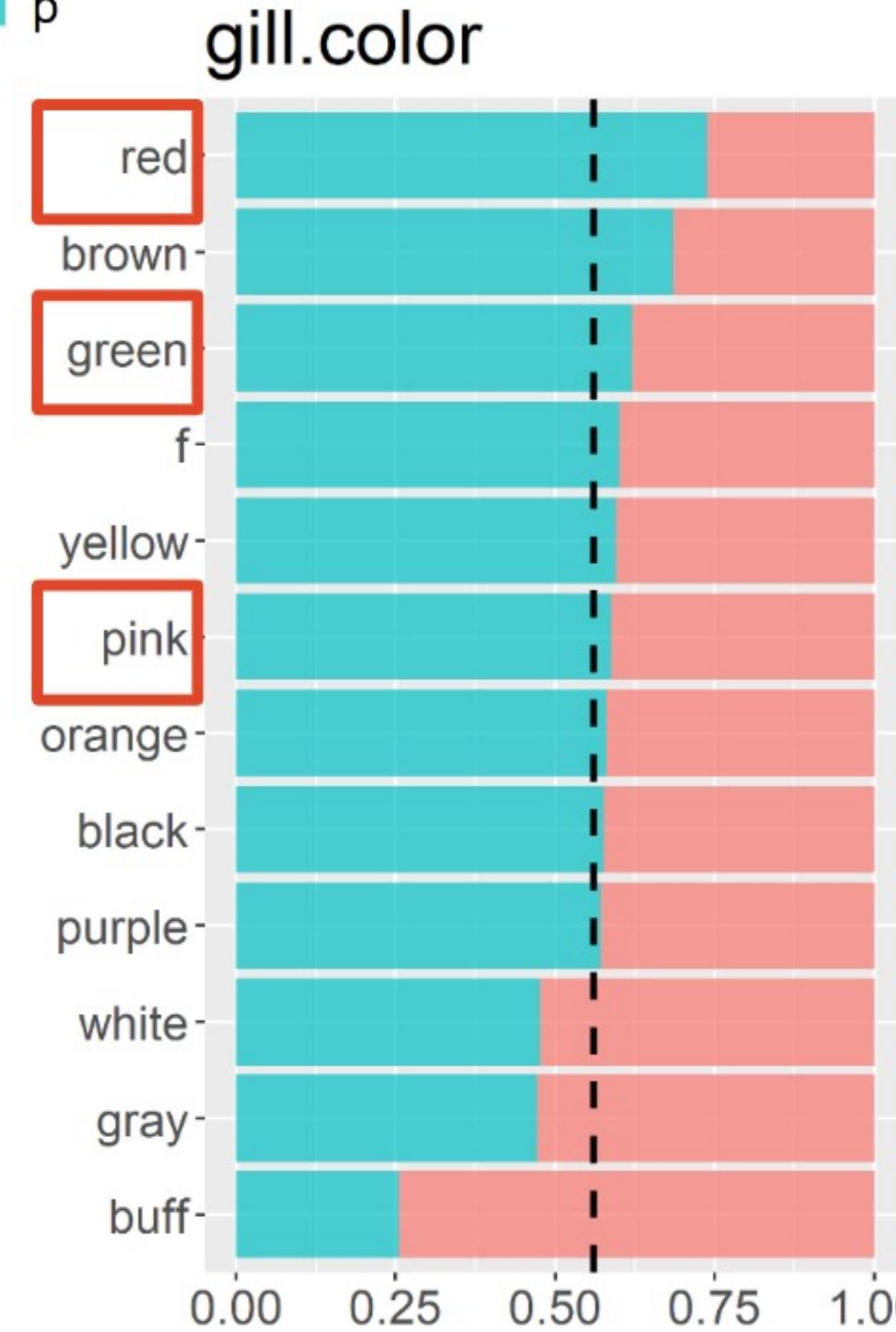
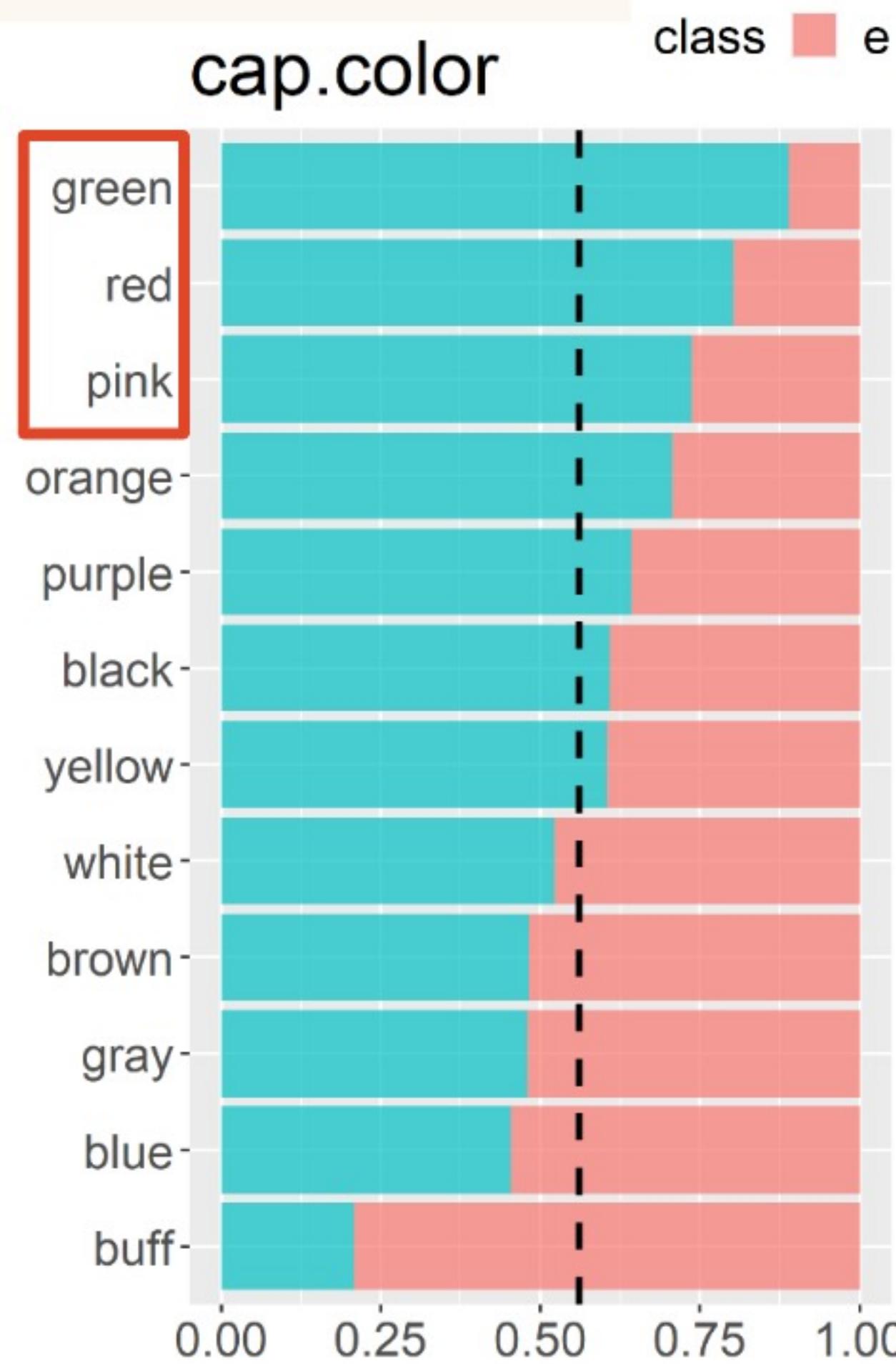
EDA : Size of Mushroom



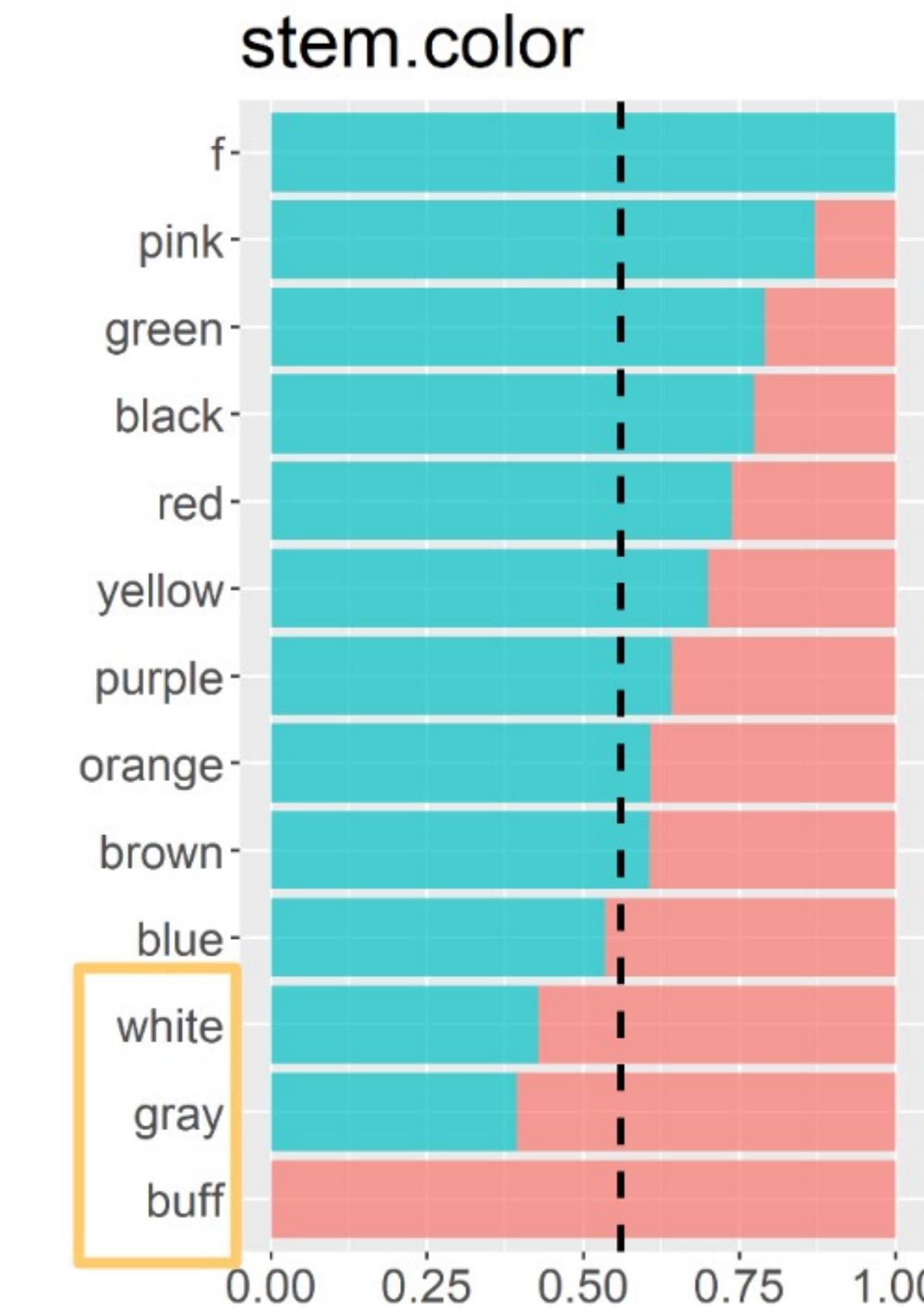
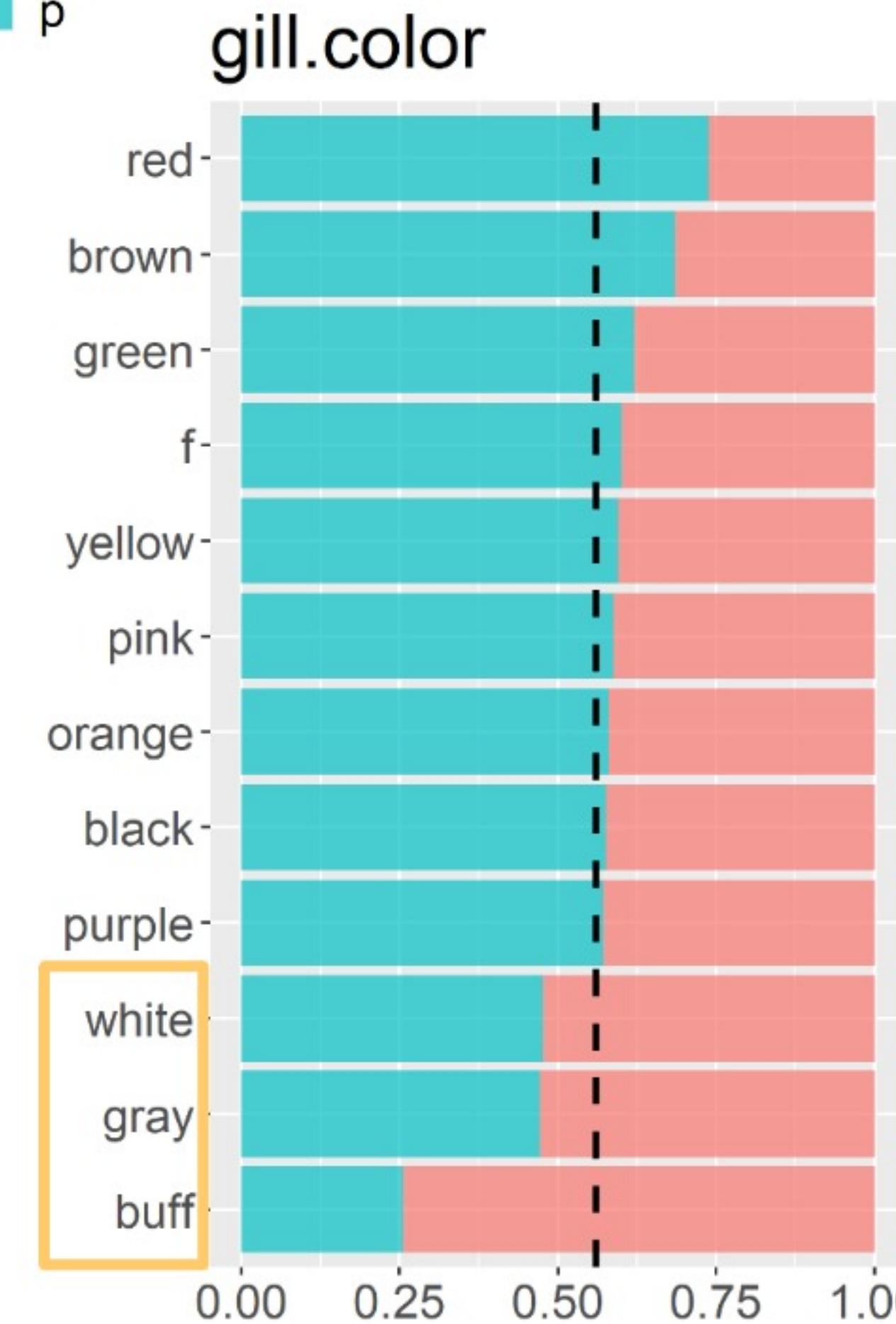
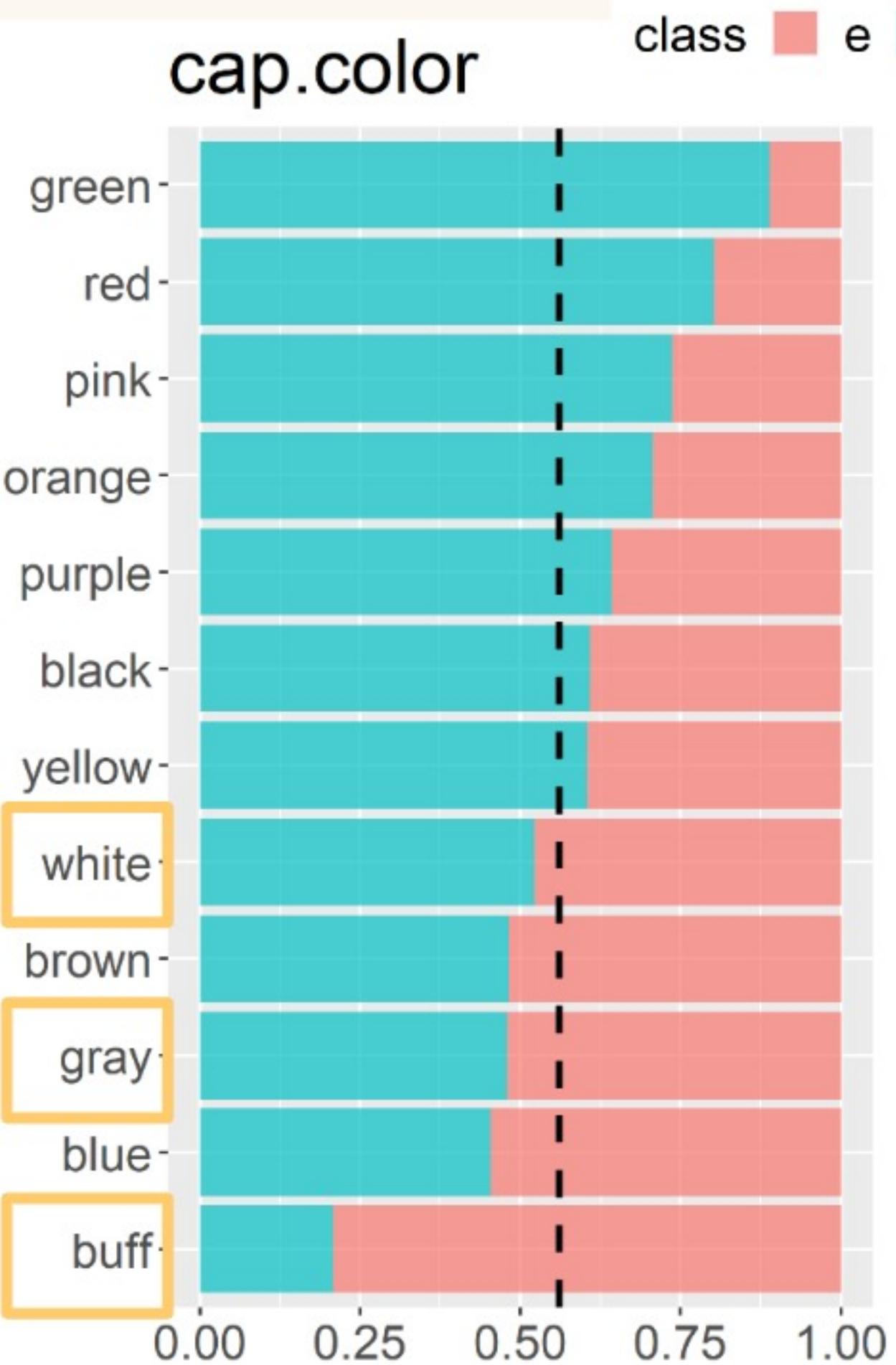
EDA : Color of cap/ gill/ stem



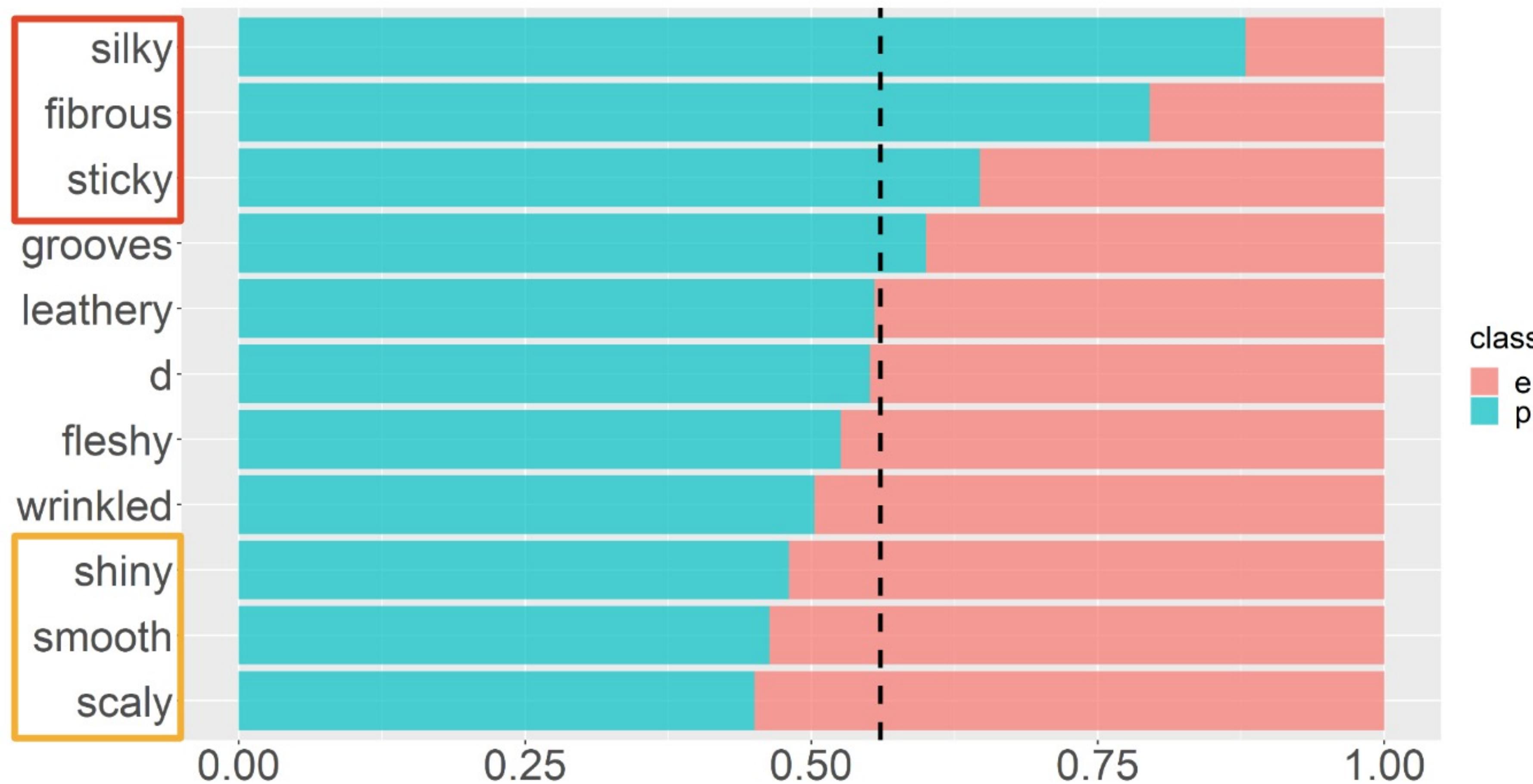
EDA : Color of cap/ gill/ stem



EDA : Color of cap/ gill/ stem



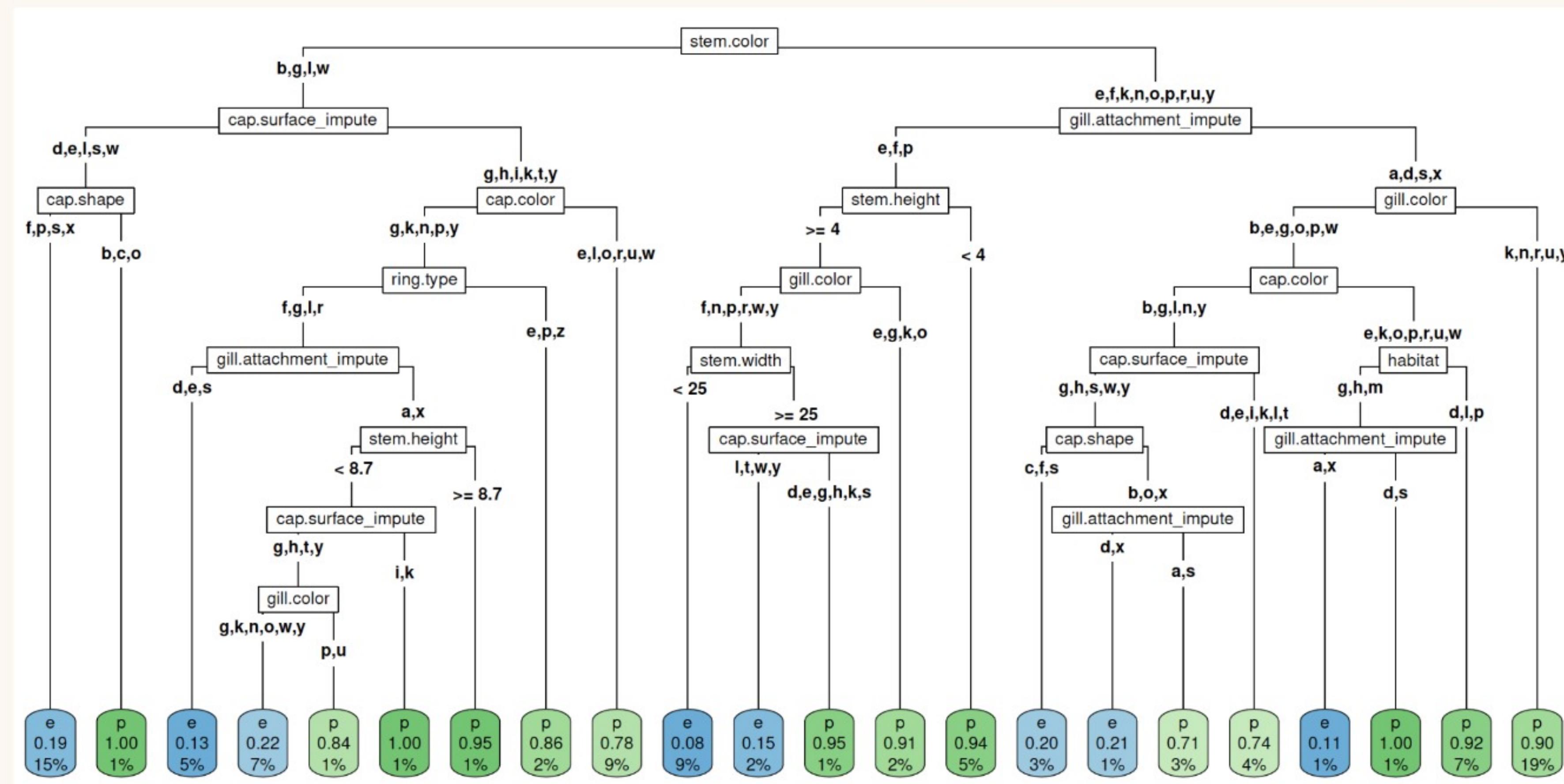
EDA : Surface of cap



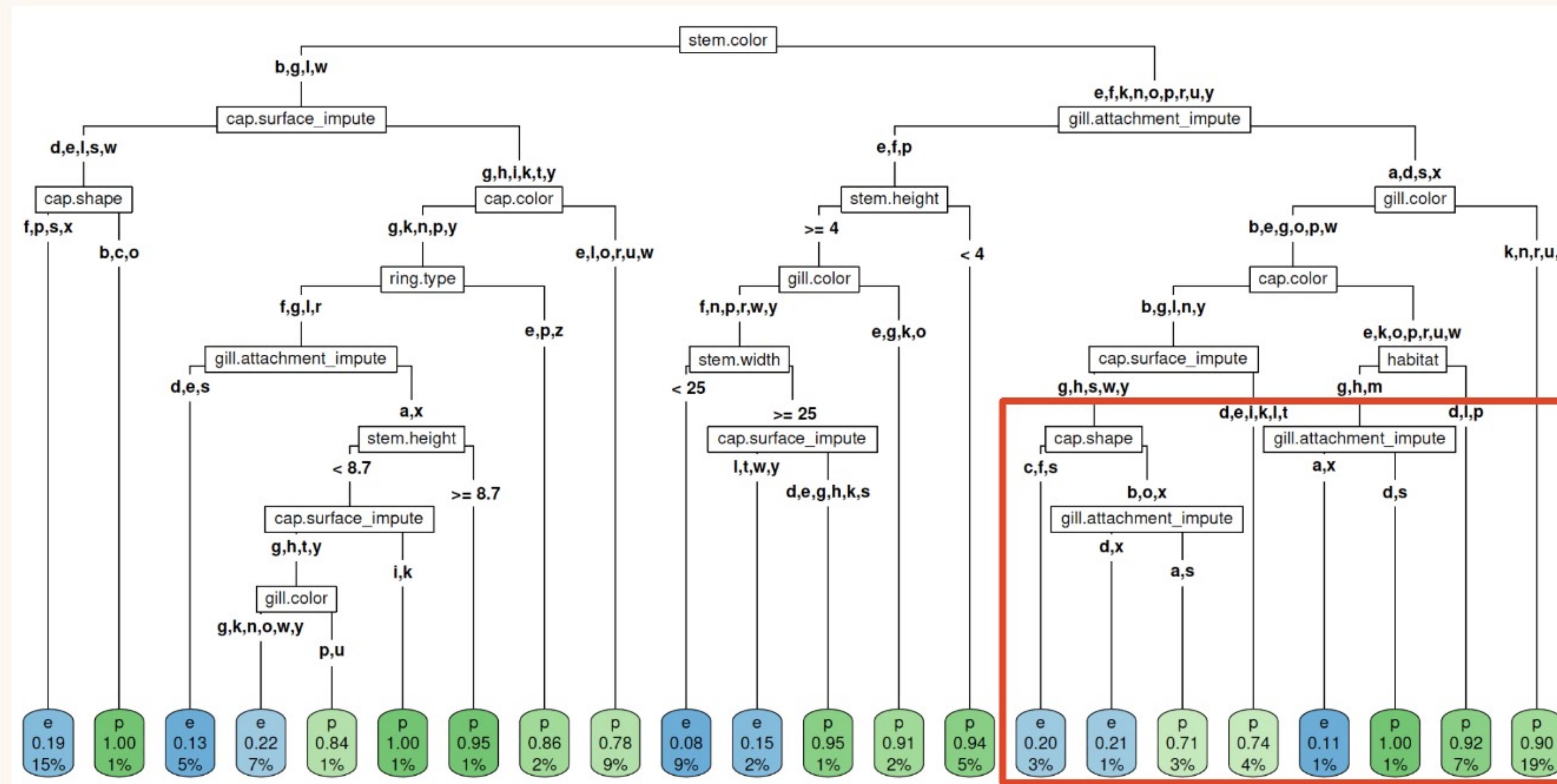
Model Performance

Model	Logistic	LDA	NB	Tree	RF	Boosting	SVM
ACC	0.7752	0.7702	0.6632	0.8513	0.9991	0.9279	0.9992
AUC	0.85	0.845	0.765	0.8864	1	0.9763	1

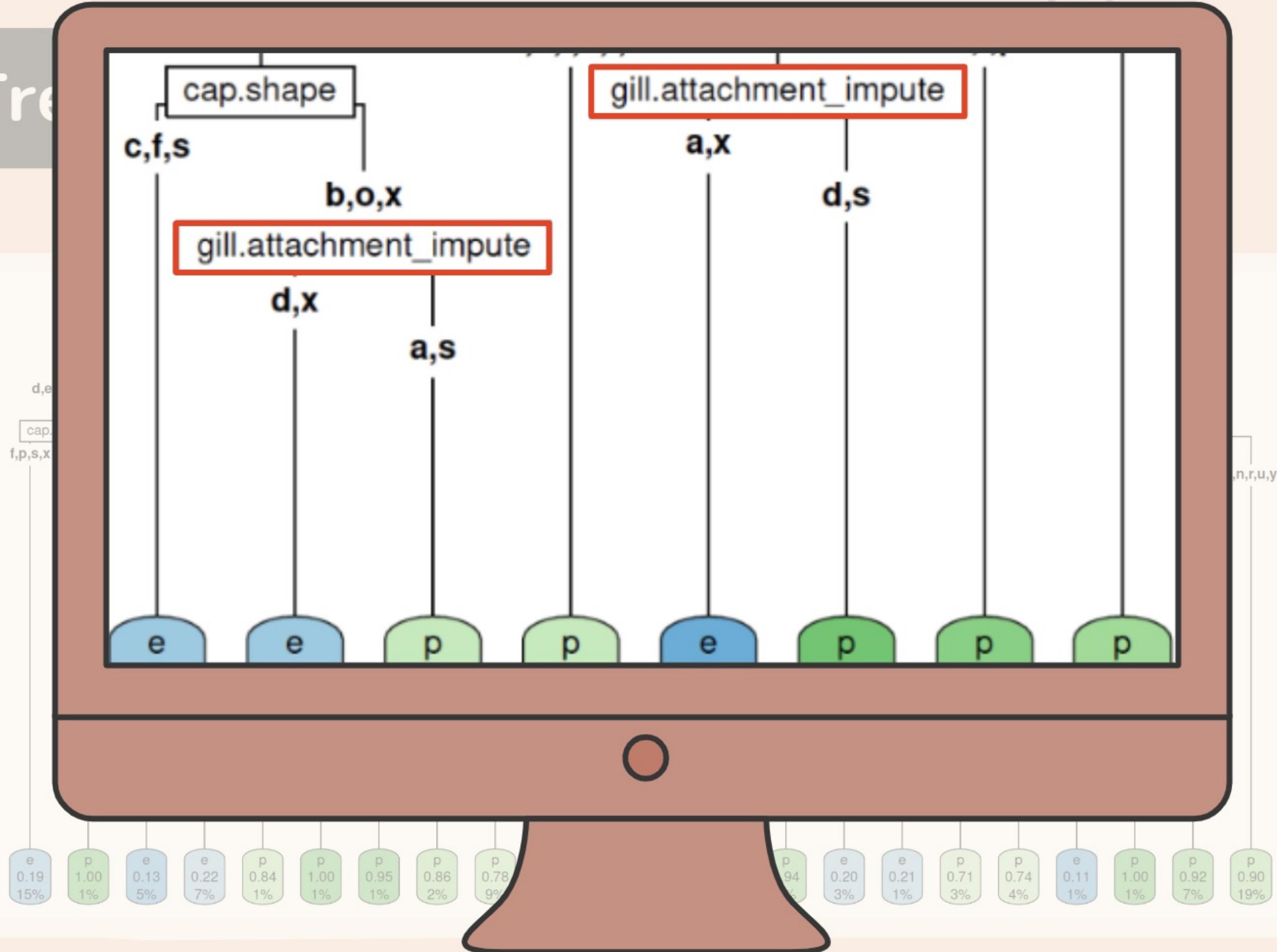
Tree-based model



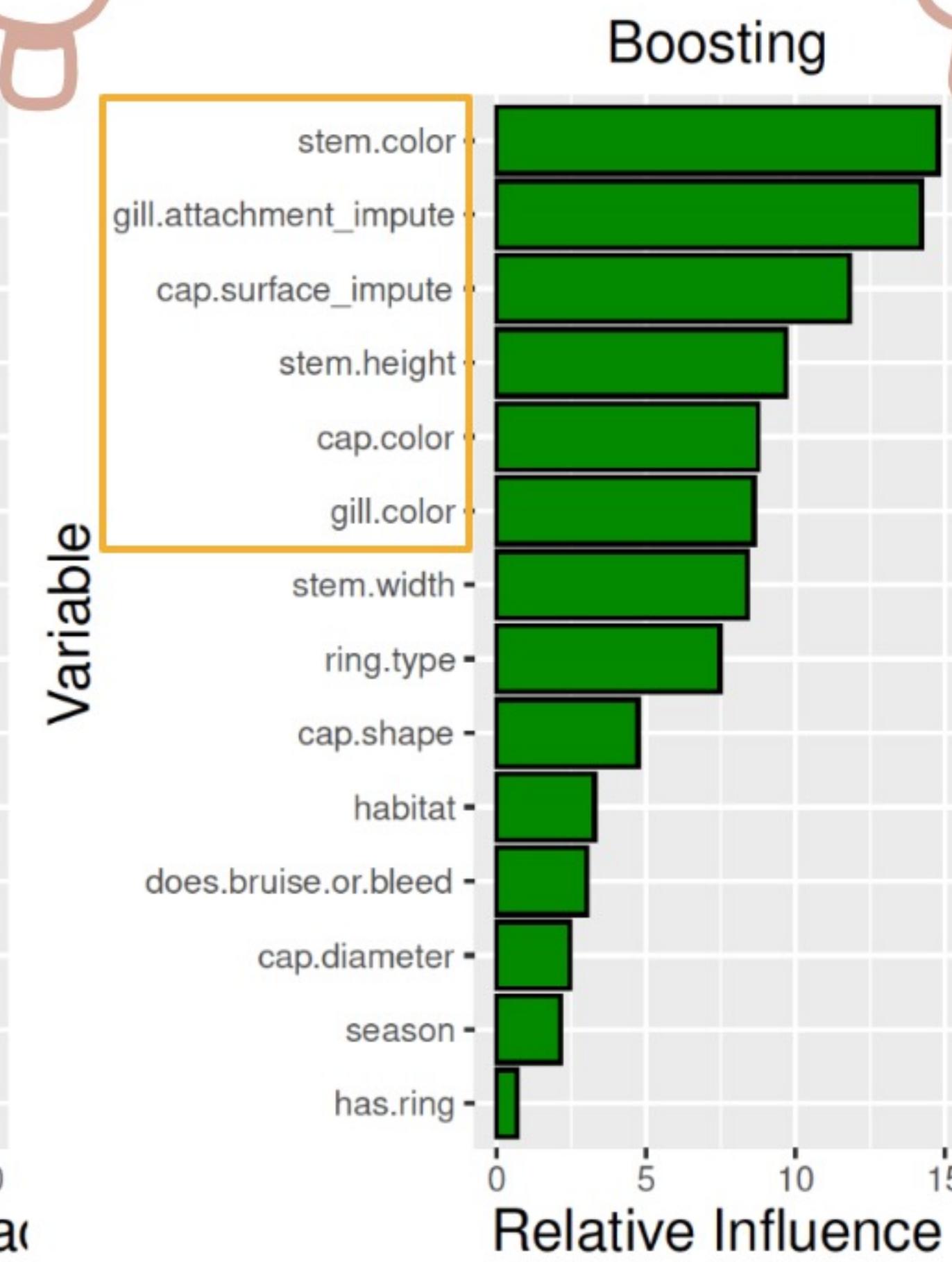
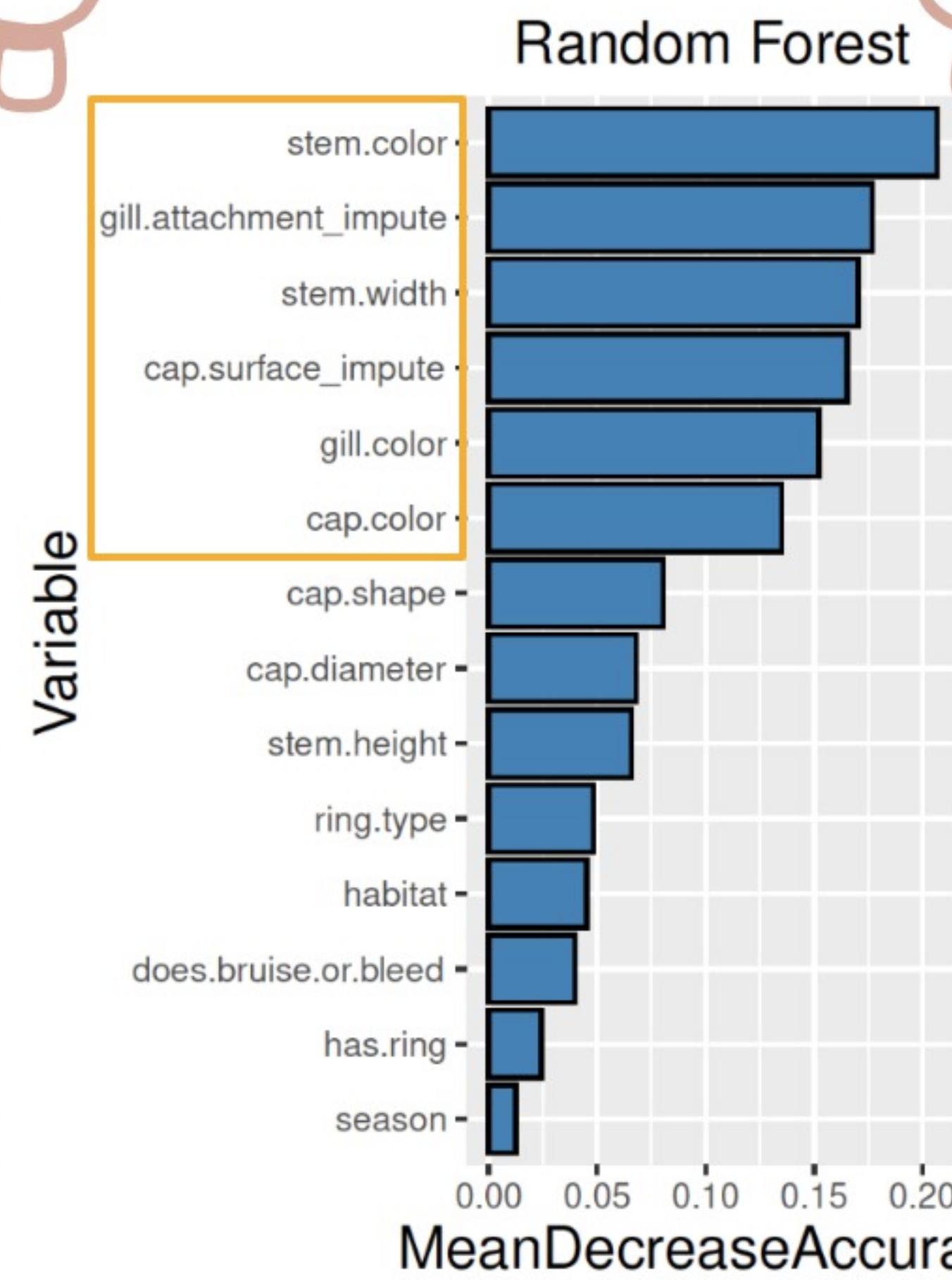
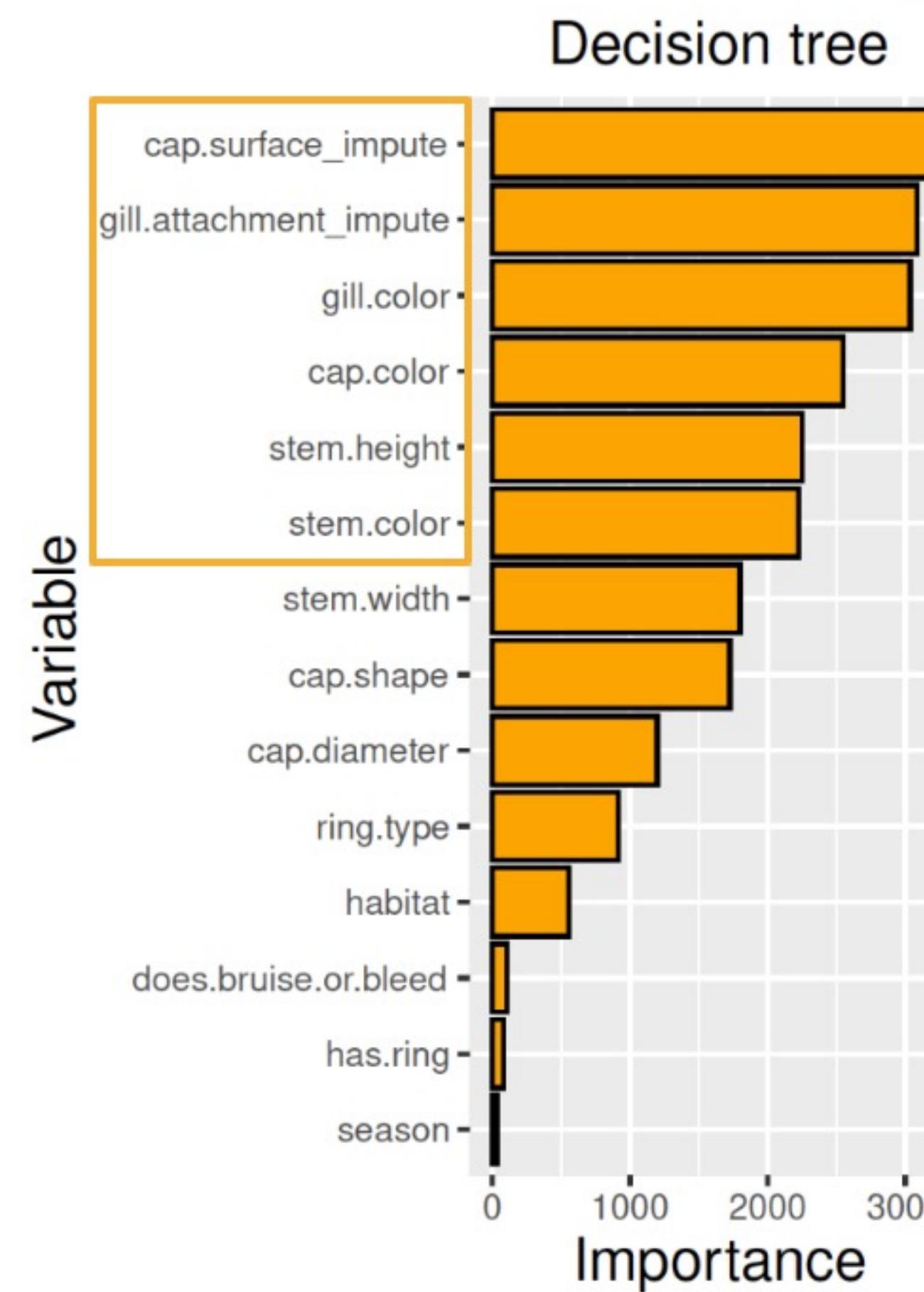
Tree-based model



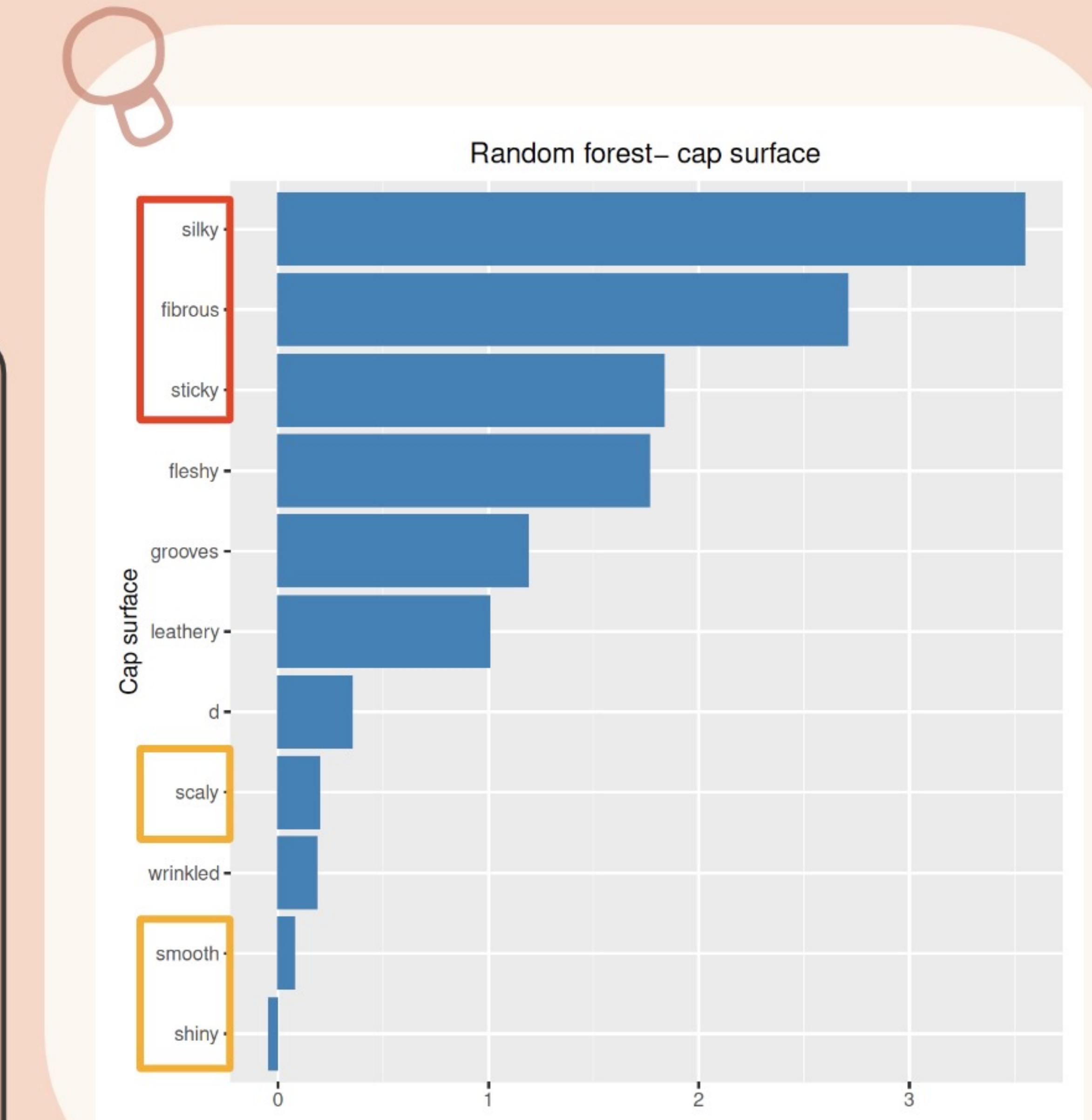
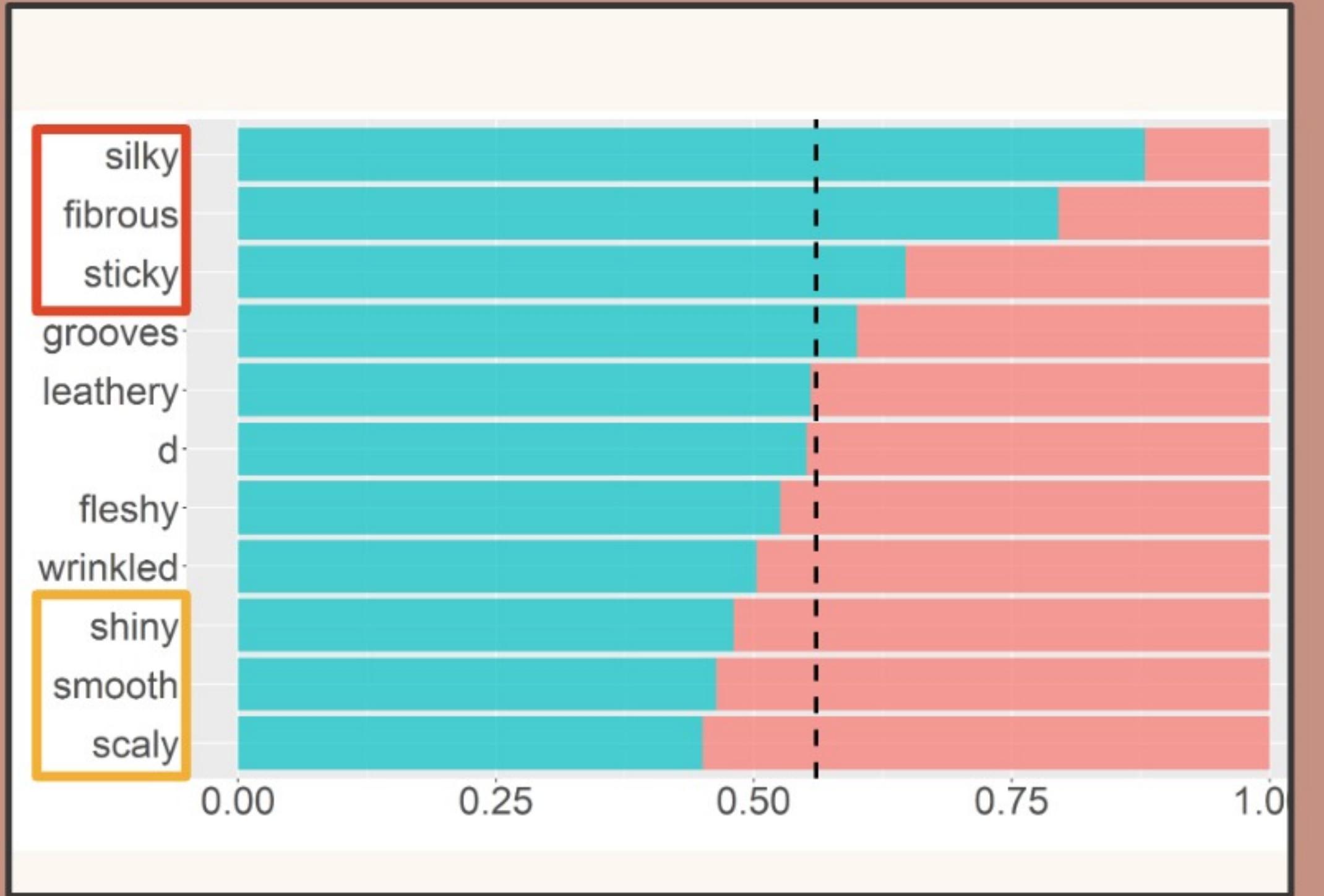
Tree



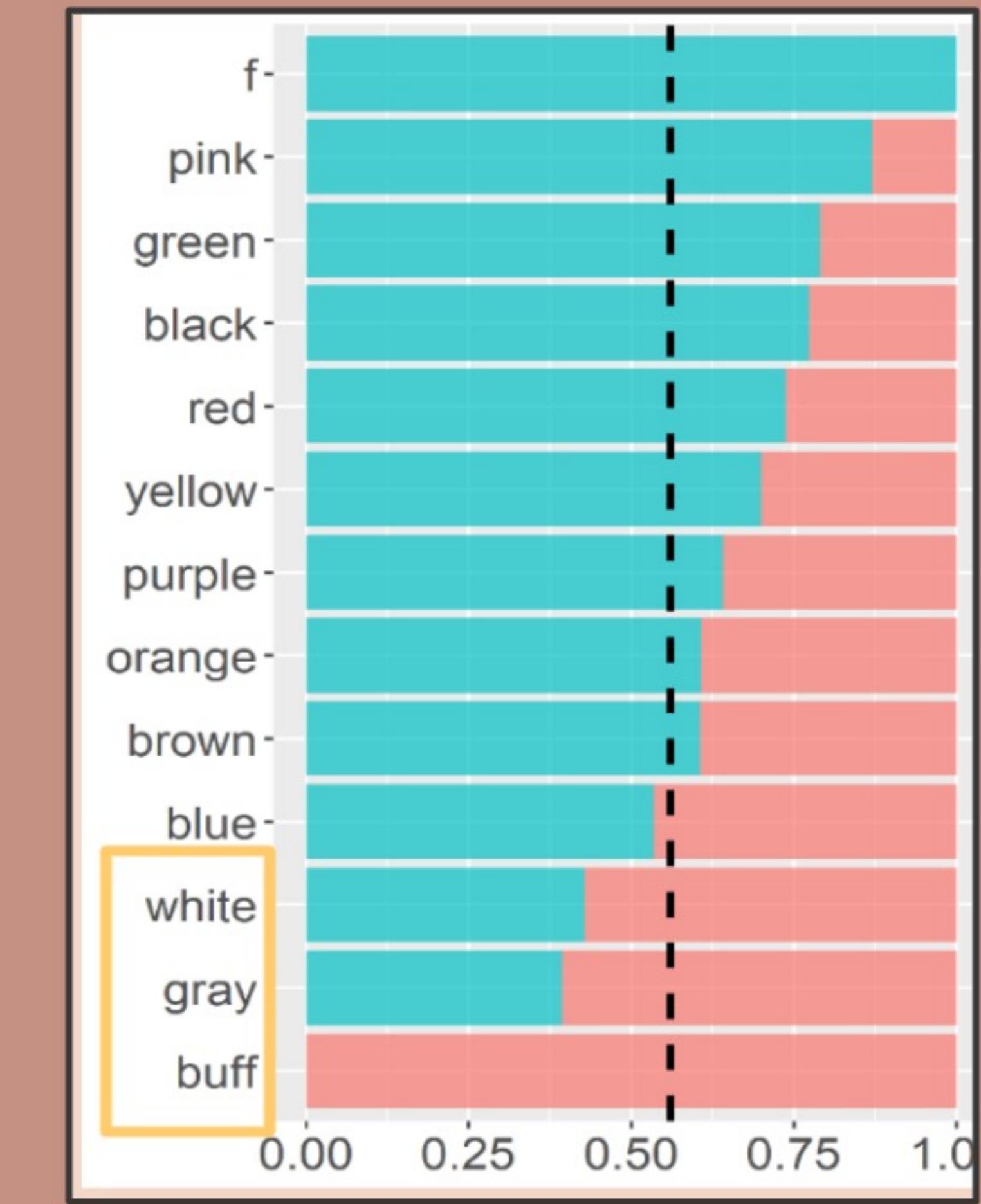
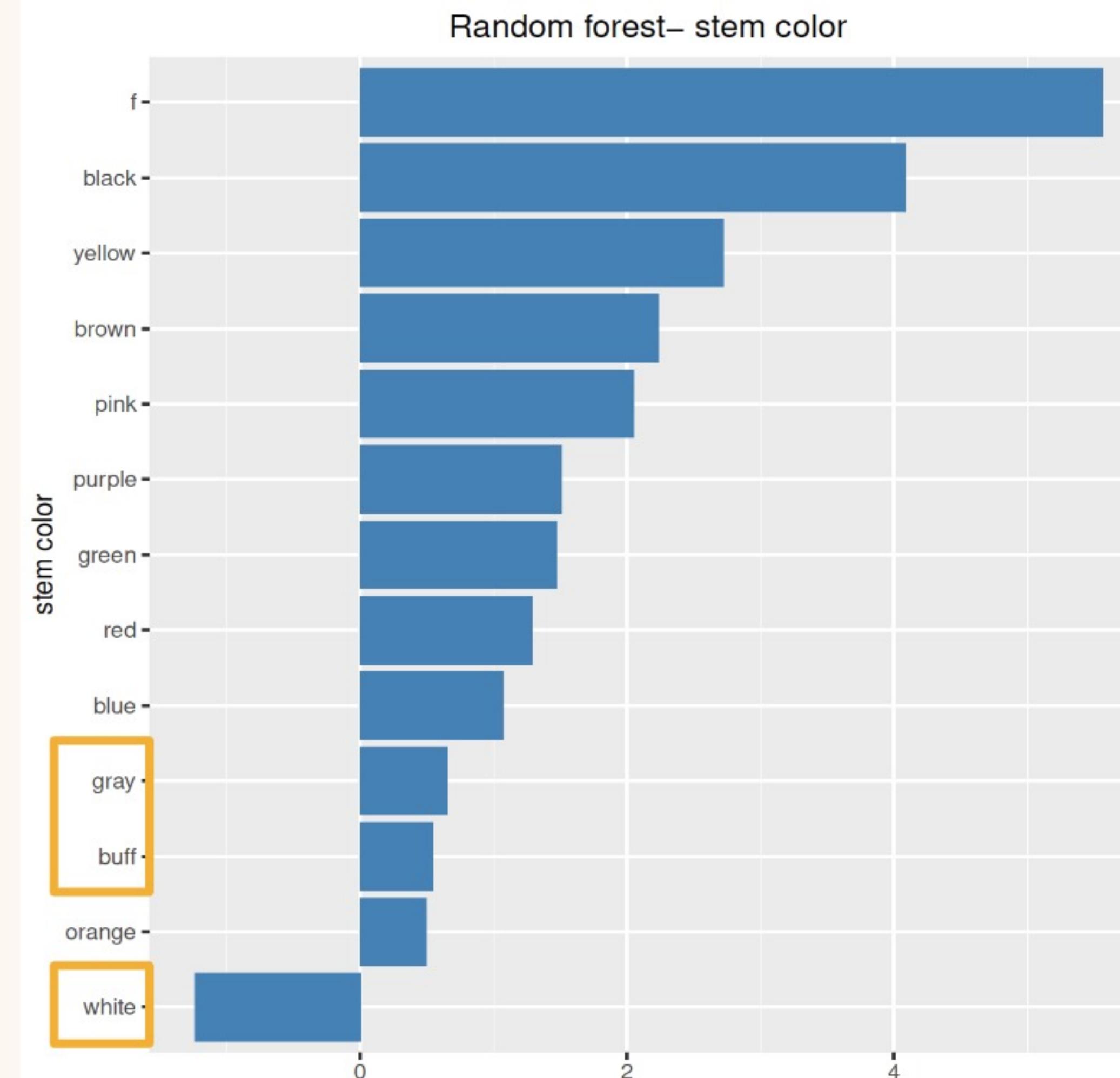
Variable Importance



PDP - cap.surface



PDP - stem.color



Predict : 毒鵝膏



season

summer, autumn



habitat woods

gill

color white
attachment free
has.ring T
ring.type zone



cap

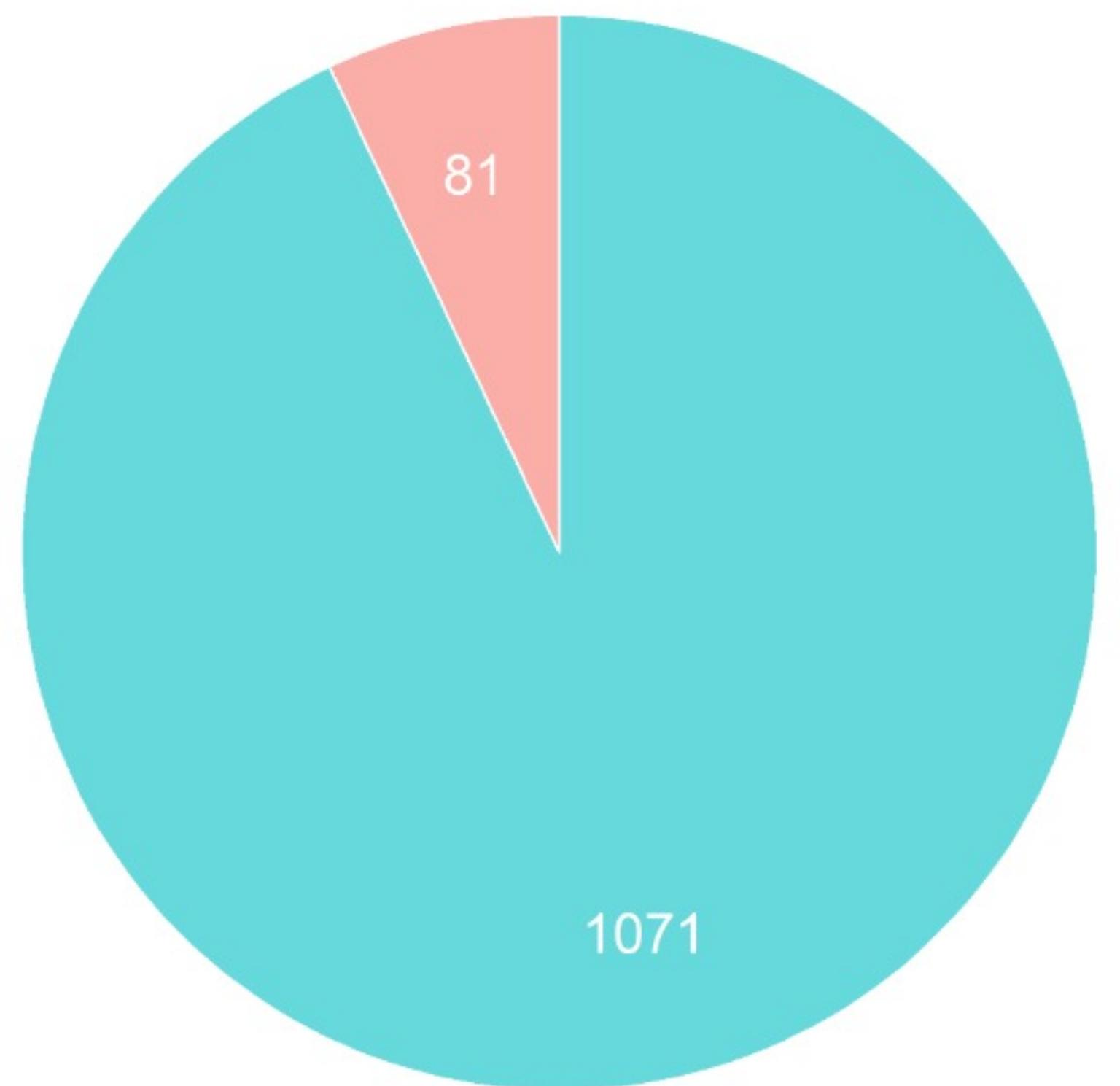
diameter 5~15 cm
surface sticky
shape convex, flat
color green
does.bruise.or.bleed T, F

stem

width 10~20 mm
height 8~15 cm
color white

Prediction Result

AVG (RF, Boosting, SVM)



毒鵝膏是有毒的喎



Conclusion



重要變數：

顏色、大小、表面質地、菌褶附著方式



蘑菇是否有毒受各變數之間的交互作用影響

只從 marginal probability (main effect) 做判斷並不精準



在較複雜的 tree-based 模型下

區分蘑菇是否有毒的準確率高達 90% 以上



邱繼賢 EDA

廖偉傑 建構模型

邱宛如 建構模型、簡報設計

data source :

<https://www.kaggle.com/datasets/devzohaib/mushroom-edibility-classification>