EDA: GAMs

Virgilio Gonzenbach

1/13/2021

Objective

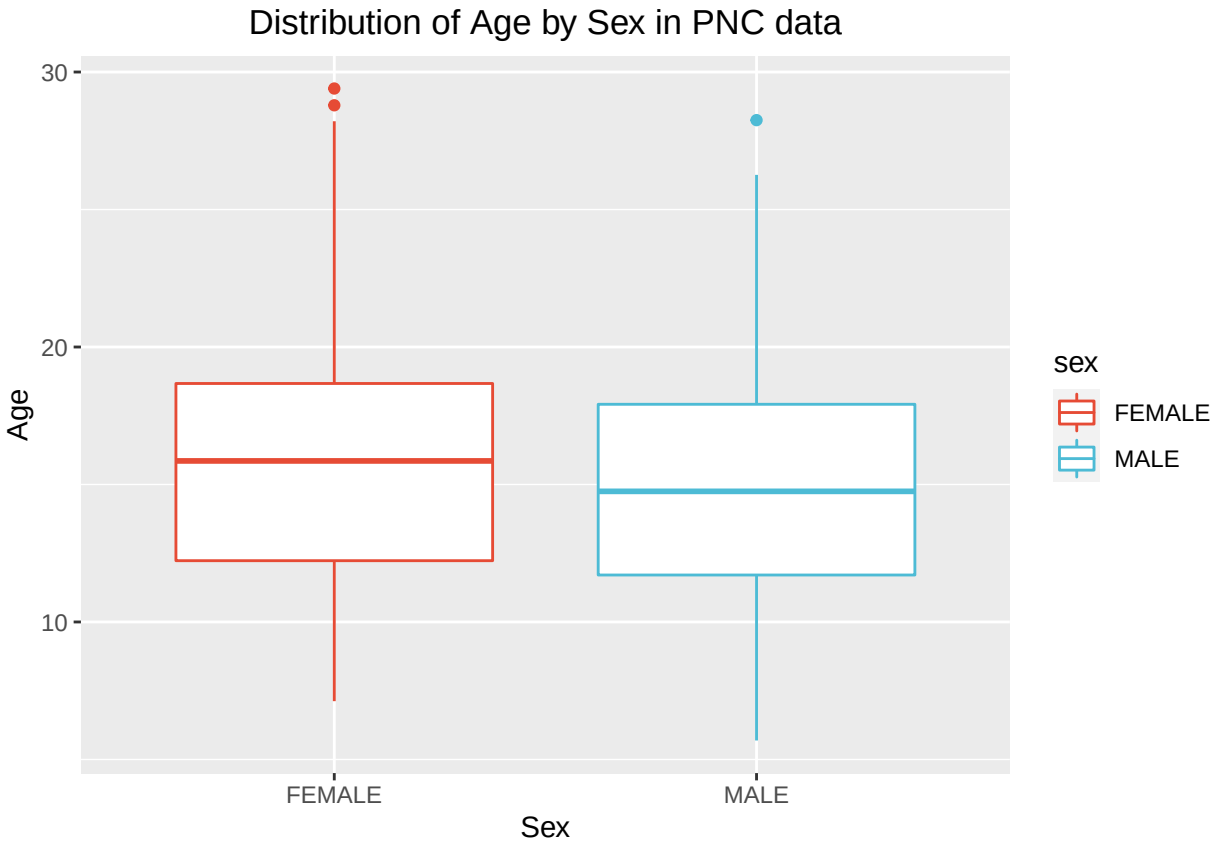
To inform the potential usage of a nonlinear method of harmonization (ComBat-GAM) by:

1. Visualizing nonlinear effects of age on various volumetric regions of interest (ROIs),
2. Modeling these nonlinear effects using Generative Additive Models (GAMs), and
3. Making comparisons between GAMs and their linear counter parts.

All modeling and visualization carried out in PNC data.

Note: All GAMs constrained to K = 3 and FX = TRUE, where K denotes the number of basis functions and FX indicates the use of fixed degrees of freedom.

Age Descriptives



Full dataset:

	n	mean	sd	median	min	max
age	1185	15.12	3.742	15.33	8.083	23.08

Males:

	n	mean	sd	median	min	max
age	555	14.78	3.717	14.67	8.167	22.92

Females:

	n	mean	sd	median	min	max
age	630	15.42	3.741	15.75	8.083	23.08

Comparison of GAMs vs Linear models across all 145 ROIs

-Model 1: ROI = sex + age + age*sex
-Model 2: ROI = sex + s(age, by = sex), where s() denotes the smooth function(s) fitted separately within each sex.

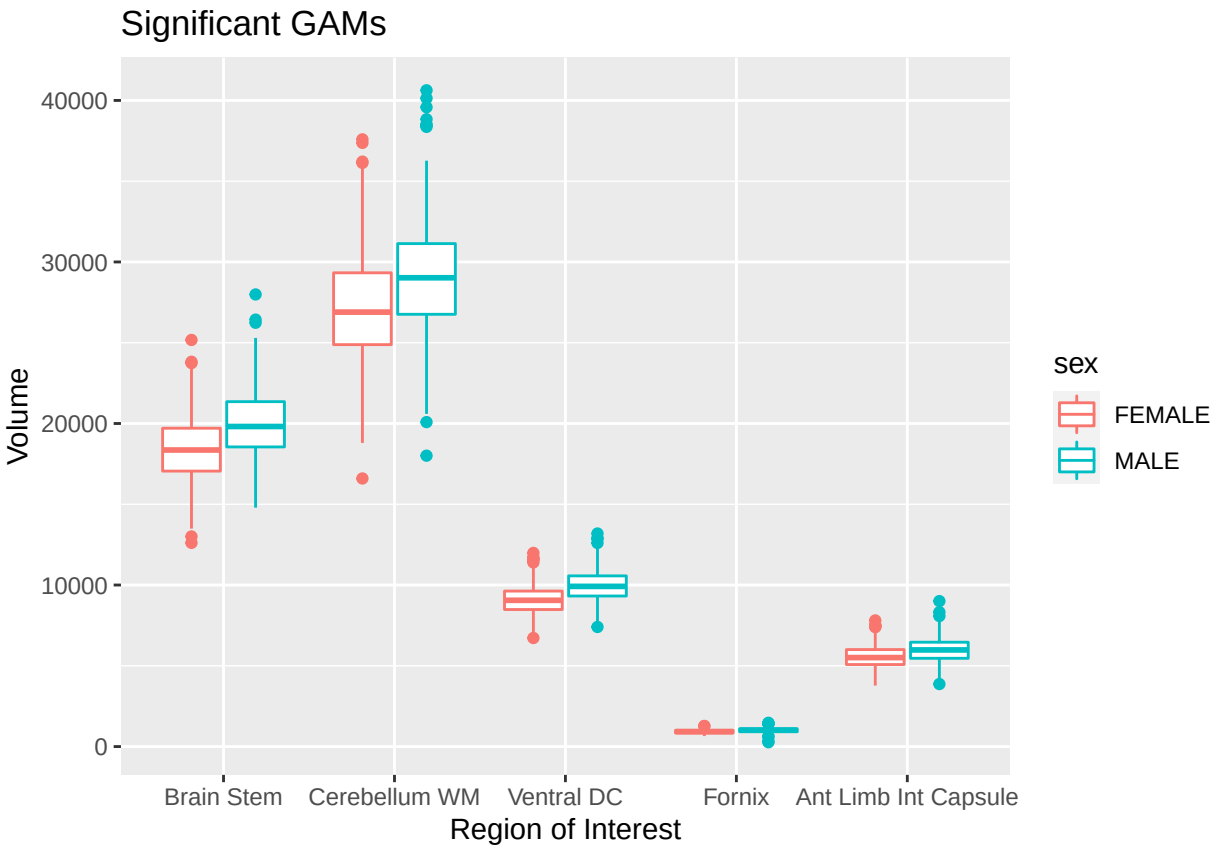
Number of regions for which models (GAM and linear model) significantly differed:

```
##    GAM != Linear (No adj.) GAM != Linear (Bonferroni)
##                         23                          3
##        GAM != Linear (FDR)
##                          9
```

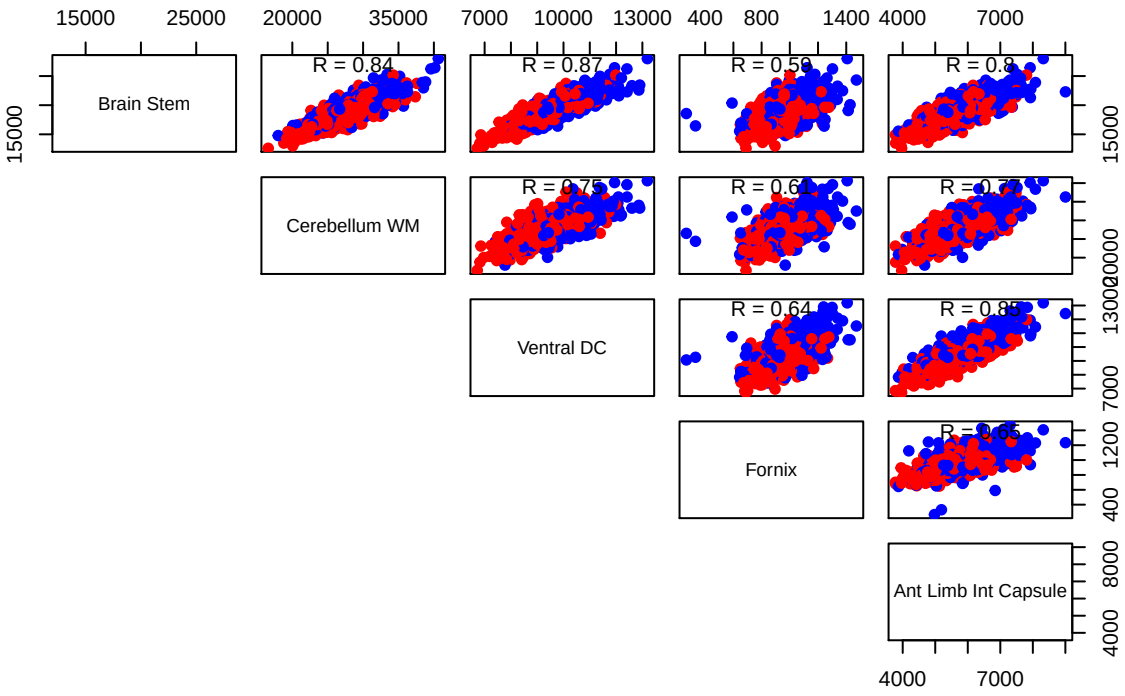
ROIs where GAM differed from linear model (adjusted for False Discovery Rate):

ROI_INDEX	ROI_NAME	HEMISPHERE	TISSUE_SEG
35	Brain Stem	Both	NONE
40	Right Cerebellum White Matter	Right	WM
41	Left Cerebellum White Matter	Left	WM
61	Right Ventral DC	Right	WM
62	Left Ventral DC	Left	WM
89	fornix right	Right	WM
90	fornix left	Left	WM
91	anterior limb of internal capsule right	Right	WM
92	anterior limb of internal capsule left	Left	WM

Regions of Interest (Significant)

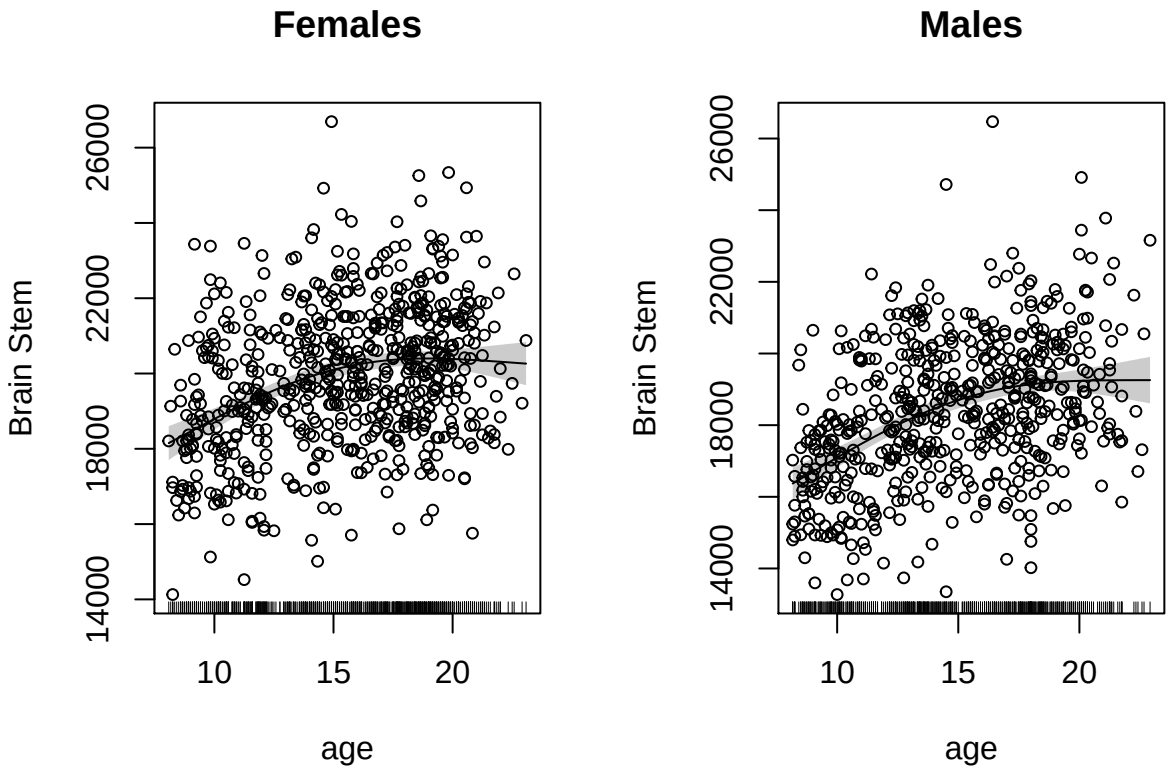



```
## NULL
```


Pairwise correlations of significant ROIs (GAM)
Brain Stem
Cerebellum WM
Ventral DC
Fornix
Ant Limb Int Capsule
R = 0.84
R = 0.87
R = 0.59
R = 0.8
R = 0.75
R = 0.64
R = 0.85
15000
25000
20000
35000
7000
10000
13000
400
800
1400
4000
7000
1200

Brain Stem

Brain Stem: Plot GAM



- Males: $R^2 = 0.178$
- Females: $R^2 = 0.112$

Brain Stem: Linear Models by Sex

Table 5:

	Dependent variable:	
	Brain Stem Volume	
	Males	Females
	(1)	(2)
Age	0.410*** (0.334, 0.485)	0.283*** (0.207, 0.358)
Age (Squared)	−0.131*** (−0.207, −0.056)	−0.137*** (−0.213, −0.062)
Constant	0.000 (−0.075, 0.075)	0.000 (−0.074, 0.074)
R^2	0.182	0.116
Note:		*p<0.05; **p<0.01; ***p<0.001
		Stargazer

Brain Stem: GAM summary

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## Brain.Stem ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18317.03      72.28  253.43   <2e-16 ***
## sexMALE      1630.40     105.67   15.43   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df     F p-value
## s(age):sexFEMALE   2      2 39.73  <2e-16 ***
## s(age):sexMALE     2      2 62.82  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.256   Deviance explained = 25.9%
## -REML =  10532  Scale est. = 3.27e+06  n = 1185
```

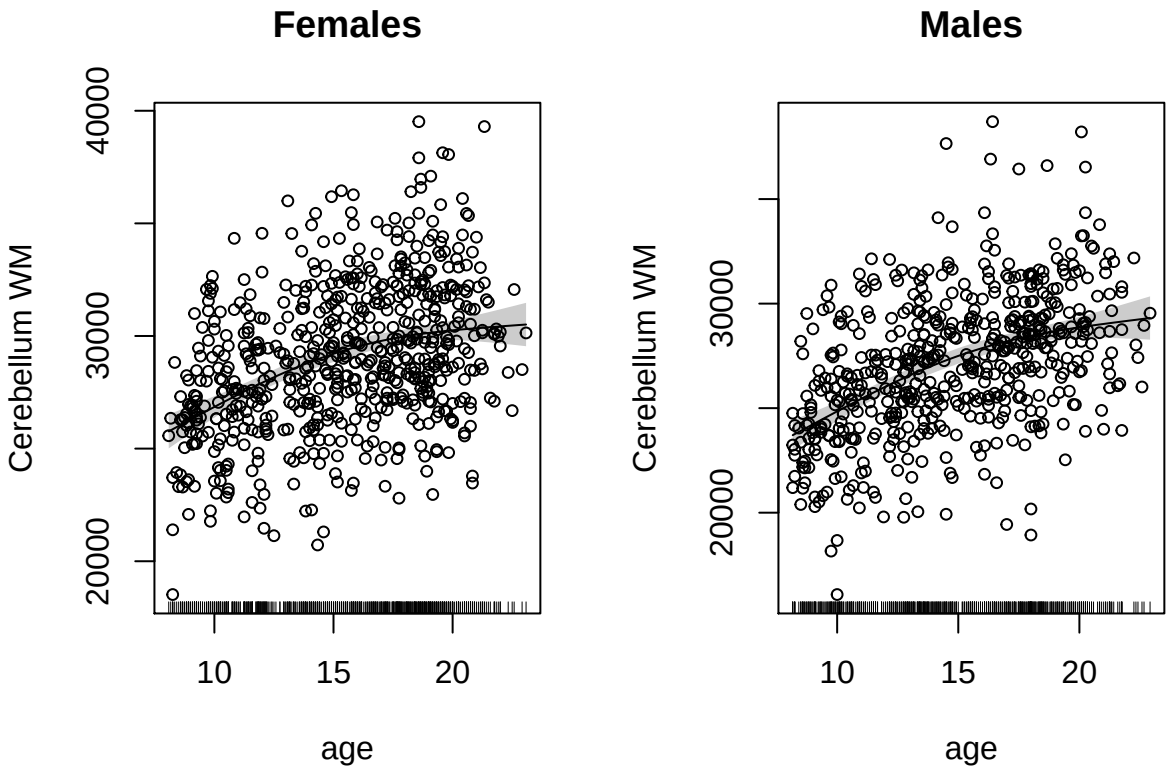
Compare with R-squared of the equivalent linear model (ICV ~ sex + age + age x sex): 0.244

Compare GAM and linear model directly:

```
## Analysis of Variance Table
##
## Model 1: Brain.Stem ~ scale(age) + sex * scale(age) + sex
## Model 2: Brain.Stem ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##   Res.Df        RSS Df Sum of Sq      F    Pr(>F)
## 1   1181 3933341148
## 2   1179 3855353802  2  77987346 11.925 7.463e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cerebellum White Matter

Cerebellum WM: Plot GAM



Cerebellum WM: Linear Models by Sex

Table 6:

	Dependent variable:	
	Cerebellum WM	
	Males	Females
	(1)	(2)
Age	0.464*** (0.390, 0.537)	0.365*** (0.291, 0.438)
Age (Squared)	−0.099** (−0.173, −0.026)	−0.106** (−0.179, −0.032)
Constant	0.000 (−0.074, 0.074)	0.000 (−0.072, 0.072)
R^2	0.222	0.161
Note:		*p<0.05; **p<0.01; ***p<0.001
		Stargazer

R-squared of GAMs:
- Males: $R^2 = 0.219$
- Females: $R^2 = 0.157$

Cerebellum WM: GAM summary

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## Cerebellum.WM ~ s(age, by = sex, bs = "tp", k = K, fx = FX) +
##     sex
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26961.2      118.6  227.35   <2e-16 ***
## sexMALE       2151.8      173.4   12.41   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df     F p-value
## s(age):sexFEMALE   2      2 61.15  <2e-16 ***
## s(age):sexMALE     2      2 76.27  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =   0.25   Deviance explained = 25.3%
## -REML =  11116  Scale est. = 8.803e+06  n = 1185
```

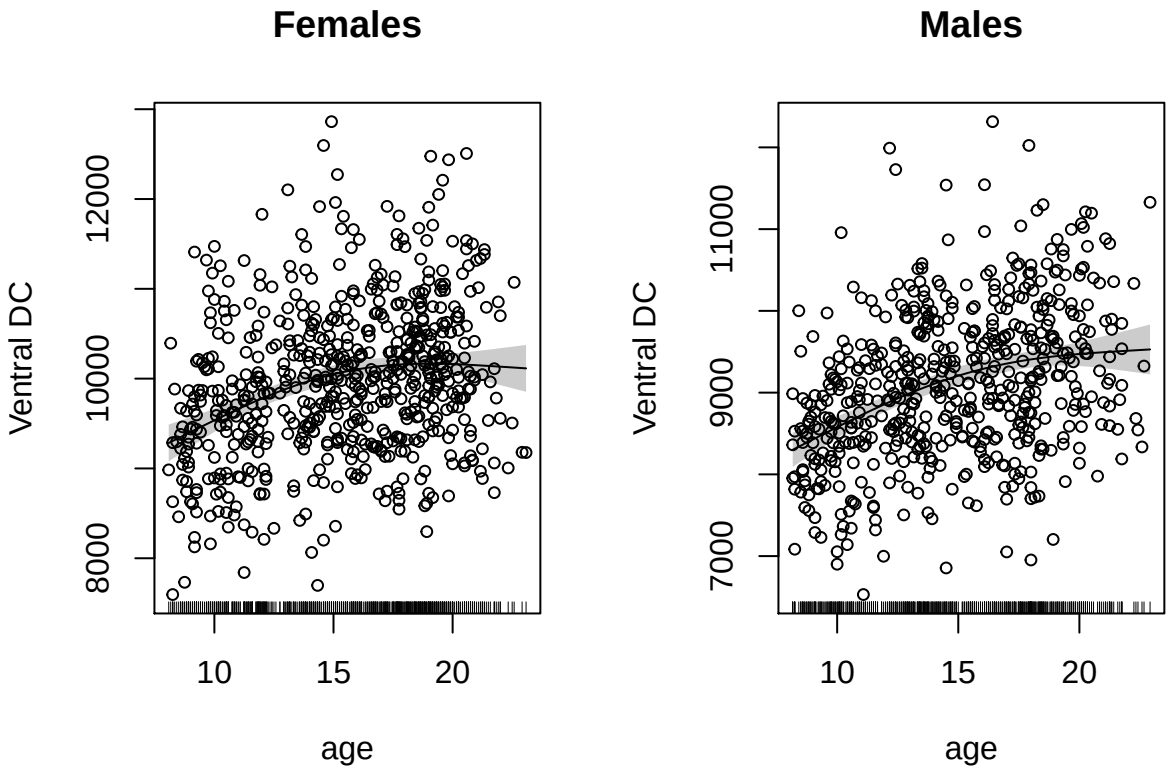
Compare with R-squared of the equivalent linear model (Cerebellum.WM ~ sex + age + age x sex): 0.244

Compare GAM and linear model directly:

```
## Analysis of Variance Table
##
## Model 1: Cerebellum.WM ~ scale(age) + sex * scale(age) + sex
## Model 2: Cerebellum.WM ~ s(age, by = sex, bs = "tp", k = K, fx = FX) +
##     sex
##   Res.Df        RSS Df Sum of Sq    F   Pr(>F)
## 1   1181 1.0501e+10
## 2   1179 1.0379e+10  2 122185165 6.94 0.001008 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ventral Diencephalon

Ventral DC: Plot GAM



Ventral DC: Linear Models by Sex

Table 7:

	Dependent variable:	
	Ventral DC	
	Males	Females
	(1)	(2)
Age	0.370*** (0.292, 0.447)	0.247*** (0.170, 0.323)
Age (Squared)	−0.099* (−0.176, −0.022)	−0.124** (−0.201, −0.048)
Constant	0.000 (−0.077, 0.077)	−0.000 (−0.075, 0.075)
R^2	0.144	0.089
Note:		*p<0.05; **p<0.01; ***p<0.001
		Stargazer

- Males: $R^2 = 0.141$
- Females: $R^2 = 0.085$

Ventral DC: GAM summary

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## Ventral.DC ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9070.02      33.46  271.09   <2e-16 ***
## sexMALE       921.18      48.92   18.83   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df     F  p-value
## s(age):sexFEMALE   2      2 28.64 7.17e-13 ***
## s(age):sexMALE     2      2 49.12  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.285   Deviance explained = 28.8%
## -REML = 9623.7  Scale est. = 7.0074e+05  n = 1185
```

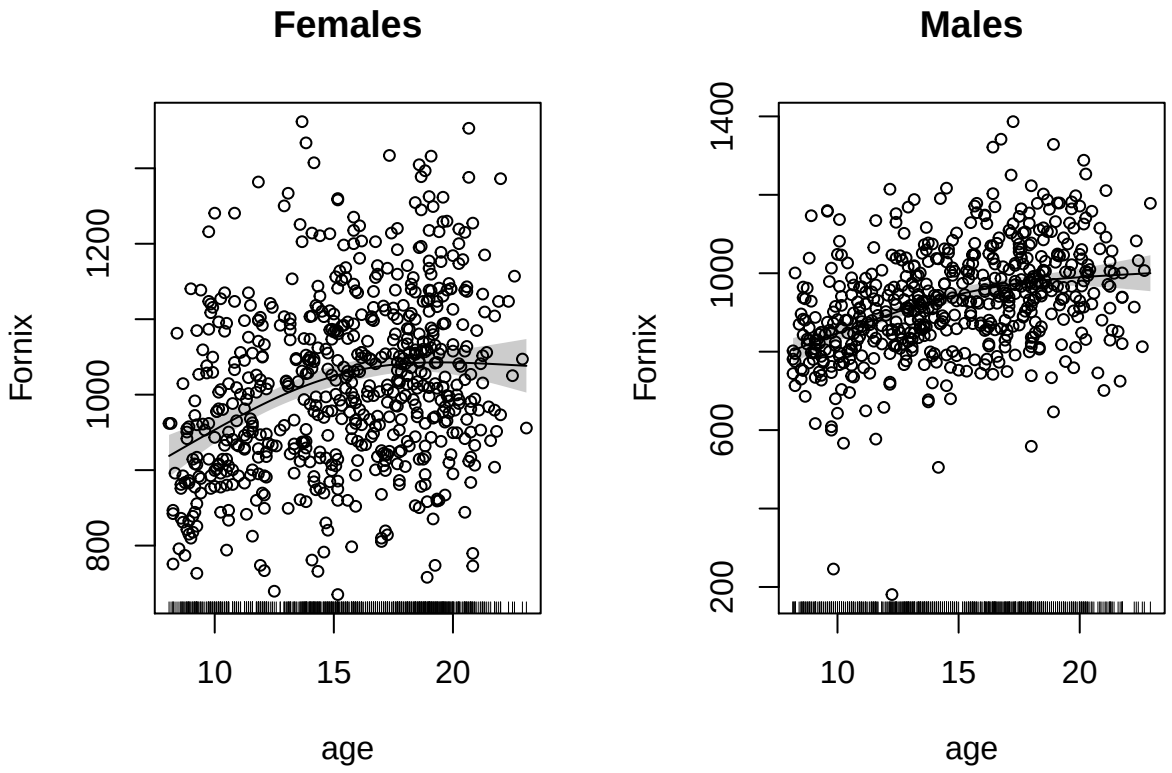
Compare with R-squared of the equivalent linear model (Ventral.DC ~ sex + age + age x sex): 0.279

Compare GAM and linear model directly:

```
## Analysis of Variance Table
##
## Model 1: Ventral.DC ~ scale(age) + sex * scale(age) + sex
## Model 2: Ventral.DC ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##   Res.Df       RSS Df Sum of Sq      F    Pr(>F)
## 1   1181 836708505
## 2   1179 826177475  2  10531031 7.5142 0.0005718 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fornix

Fornix: Plot GAM



- Males: $R^2 = 0.156$
- Females: $R^2 = 0.091$

Fornix: Linear Models by Sex

Table 8:

	Dependent variable:	
	Fornix	
	Males	Females
	(1)	(2)
Age	0.389*** (0.313, 0.466)	0.260*** (0.184, 0.337)
Age (Squared)	−0.097* (−0.174, −0.020)	−0.119** (−0.195, −0.042)
Constant	0.000 (−0.076, 0.076)	0.000 (−0.074, 0.074)
R^2	0.158	0.095
Note:		*p<0.05; **p<0.01; ***p<0.001
		Stargazer

Fornix: GAM summary

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## Fornix ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  925.194      4.790  193.14   <2e-16 ***
## sexMALE       92.724      7.004   13.24   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df     F p-value
## s(age):sexFEMALE   2      2 27.79 1.6e-12 ***
## s(age):sexMALE     2      2 60.77 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.214   Deviance explained = 21.7%
## -REML = 7332.1  Scale est. = 14365     n = 1185
```

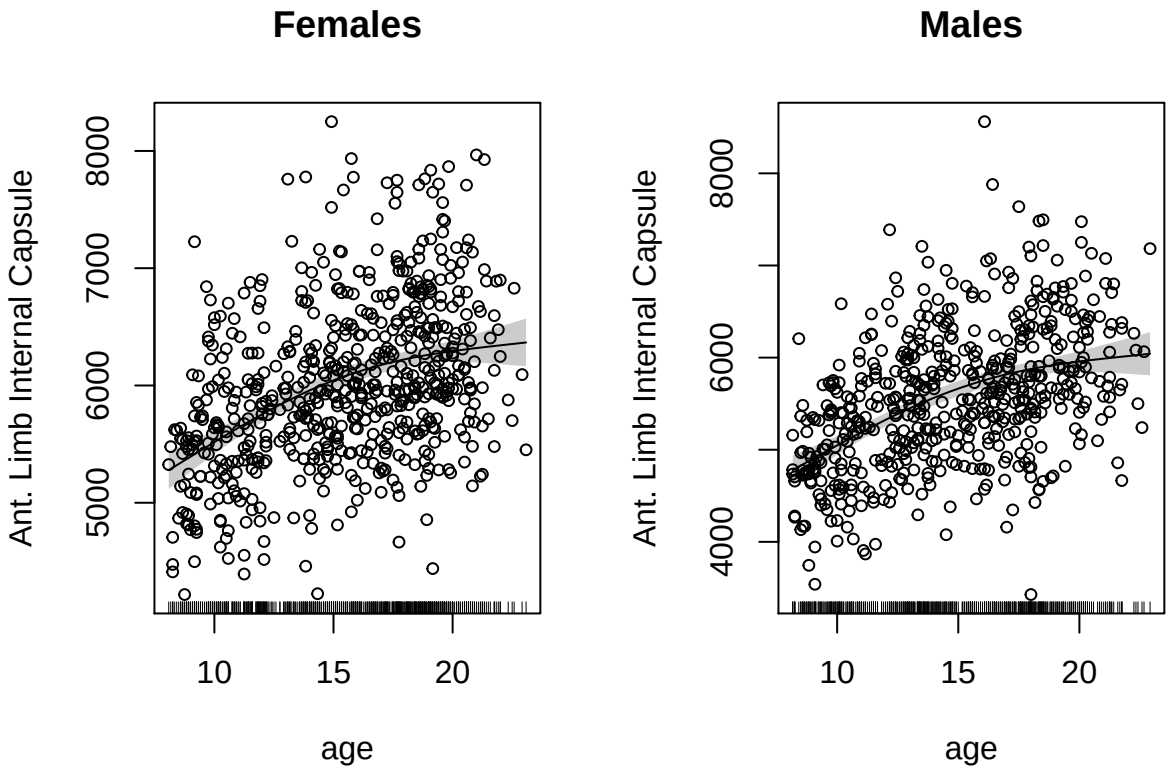
Compare with R-squared of the equivalent linear model (Fornix ~ sex + age + age x sex): 0.208

Compare GAM and linear model directly:

```
## Analysis of Variance Table
##
## Model 1: Fornix ~ scale(age) + sex * scale(age) + sex
## Model 2: Fornix ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1   1181 17147447
## 2   1179 16935969  2    211479 7.3611 0.0006652 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anterior Limb of Internal Capsule (ALIC)

ALIC: Plot GAM



- Males: $R^2 = 0.222$
- Females: $R^2 = 0.177$

ALIC: Linear Models by Sex

Table 9:

	Dependent variable:	
	Ant. Limb Internal Capsule	
	Males	Females
	(1)	(2)
Age	0.465*** (0.392, 0.539)	0.390*** (0.318, 0.463)
Age (Squared)	−0.109** (−0.183, −0.036)	−0.104** (−0.176, −0.031)
Constant	0.000 (−0.073, 0.073)	−0.000 (−0.071, 0.071)
R^2	0.225	0.180
Note:		*p<0.05; **p<0.01; ***p<0.001
		Stargazer

ALIC: GAM summary

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## Ant.Limb.Int.Capsule ~ s(age, by = sex, bs = "tp", k = K, fx = FX) +
##     sex
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5533.21      25.88  213.76   <2e-16 ***
## sexMALE       494.16      37.85   13.06   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df     F p-value
## s(age):sexFEMALE   2      2 66.10  <2e-16 ***
## s(age):sexMALE     2      2 83.19  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.267   Deviance explained =   27%
## -REML = 9321.2  Scale est. = 4.1941e+05  n = 1185
```

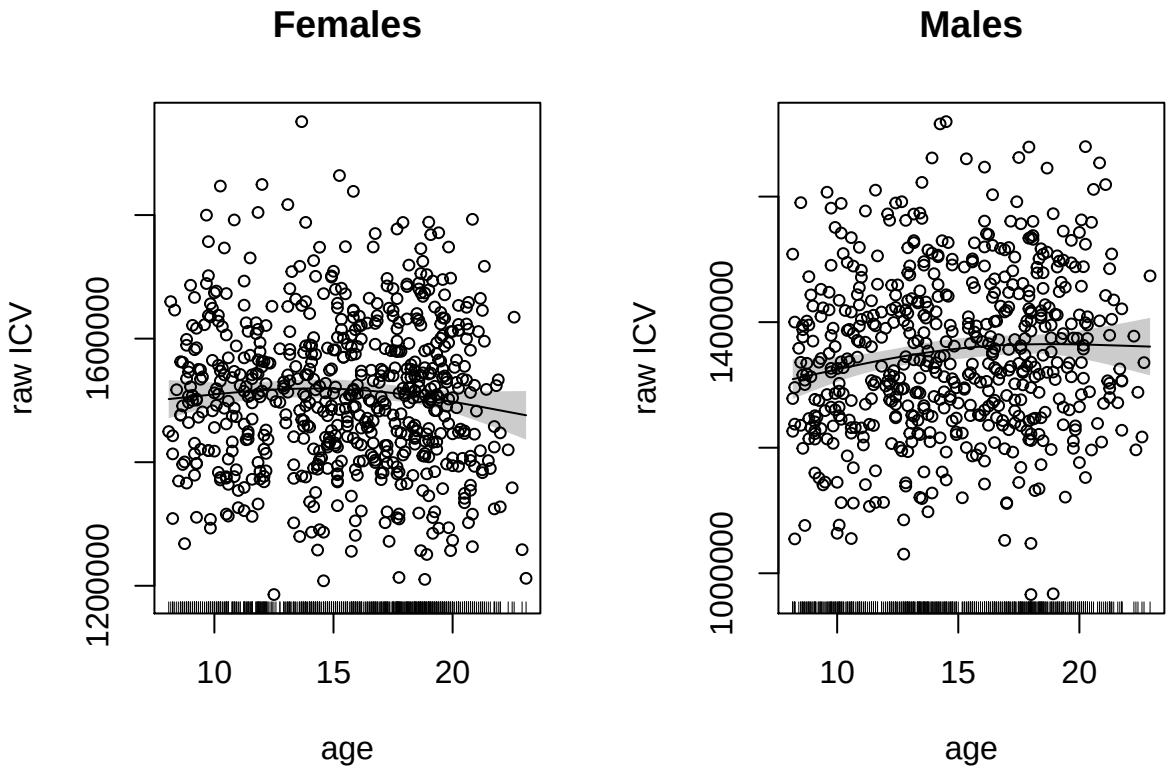
Compare with R-squared of the equivalent linear model (ALIC ~ sex + age + age x sex): 0.261

Compare GAM and linear model directly:

```
## Analysis of Variance Table
##
## Model 1: Ant.Limb.Int.Capsule ~ scale(age) + sex * scale(age) + sex
## Model 2: Ant.Limb.Int.Capsule ~ s(age, by = sex, bs = "tp", k = K, fx = FX) +
##     sex
##   Res.Df       RSS Df Sum of Sq      F    Pr(>F)
## 1   1181 500877664
## 2   1179 494488626  2   6389038 7.6166 0.0005168 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Intracranial Volume

ICV: Plot GAM



- Males: $R^2 = 0.013$
- Females: $R^2 = 0.002$

ICV: Linear Models by Sex

Table 10:

	Dependent variable:	
	Intracranial Volume	
	Males	Females
	(1)	(2)
Age	0.119** (0.037, 0.202)	−0.052 (−0.132, 0.028)
Age (Squared)	−0.053 (−0.136, 0.030)	−0.069 (−0.149, 0.011)
Constant	−0.000 (−0.083, 0.083)	0.000 (−0.078, 0.078)
R^2	0.017	0.006
Note:		*p<0.05; **p<0.01; ***p<0.001
		Stargazer

ICV: GAM summary

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## raw.ICV ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1350162       5034  268.22   <2e-16 ***
## sexMALE       160982       7360   21.87   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df     F p-value
## s(age):sexFEMALE   2      2 1.525 0.21800
## s(age):sexMALE     2      2 4.869 0.00783 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.291   Deviance explained = 29.4%
## -REML =  15535  Scale est. = 1.5862e+10  n = 1185
```

Model indicates ICV-to-Age relationship in males is approx. linear (EDF ≈ 1) and significant. In females, however, the relationship is closer to a quadratic term (EDF 2) and non-significant.

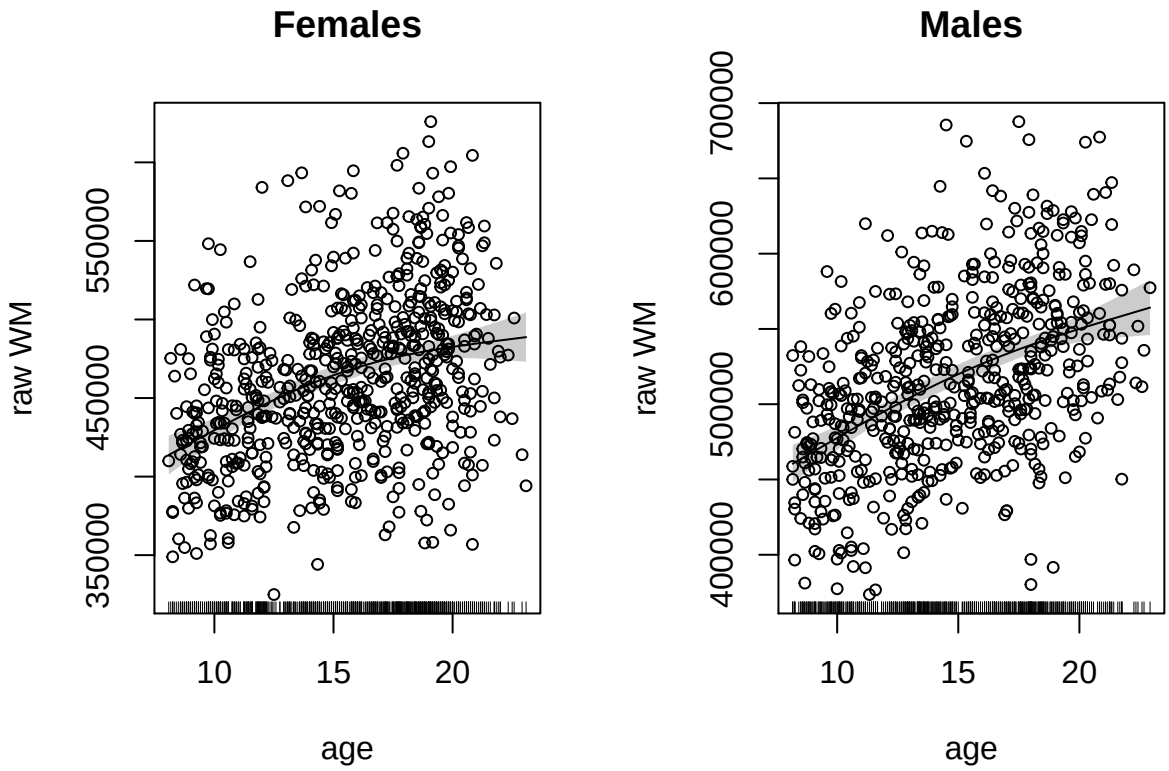
Compare with R-squared of the equivalent linear model (ICV ~ sex + age + age x sex): 0.291

Compare GAM and linear model directly:

```
## Analysis of Variance Table
##
## Model 1: raw.ICV ~ scale(age) + sex * scale(age) + sex
## Model 2: raw.ICV ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##   Res.Df        RSS Df  Sum of Sq      F Pr(>F)
## 1   1181 1.8760e+13
## 2   1179 1.8701e+13  2 5.8873e+10 1.8558 0.1568
```

White Matter

WM: Plot GAMs



- Males: $R^2 = 0.218$
- Females: $R^2 = 0.144$

WM: Linear models by sex

Table 11:

	Dependent variable:	
	White Matter Volume	
	Males	Females
	(1)	(2)
Age	0.469*** (0.395, 0.542)	0.356*** (0.282, 0.430)
Age (Squared)	−0.053 (−0.127, 0.021)	−0.086* (−0.160, −0.012)
Constant	−0.000 (−0.074, 0.074)	−0.000 (−0.072, 0.072)
R^2	0.221	0.147
Note:		*p<0.05; **p<0.01; ***p<0.001
		Stargazer

WM: GAM summary

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## WM ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   460420       2003  229.84   <2e-16 ***
## sexMALE        56987       2929   19.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df    F p-value
## s(age):sexFEMALE   2      2 51.2  <2e-16 ***
## s(age):sexMALE     2      2 82.6  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.334   Deviance explained = 33.7%
## -REML =  14448  Scale est. = 2.5119e+09  n = 1185
```

Model indicates WM-to-Age relationship is nonlinear and statistically significant in both males and females. WM increases with age; but in females, growth stops at 20 years of age.

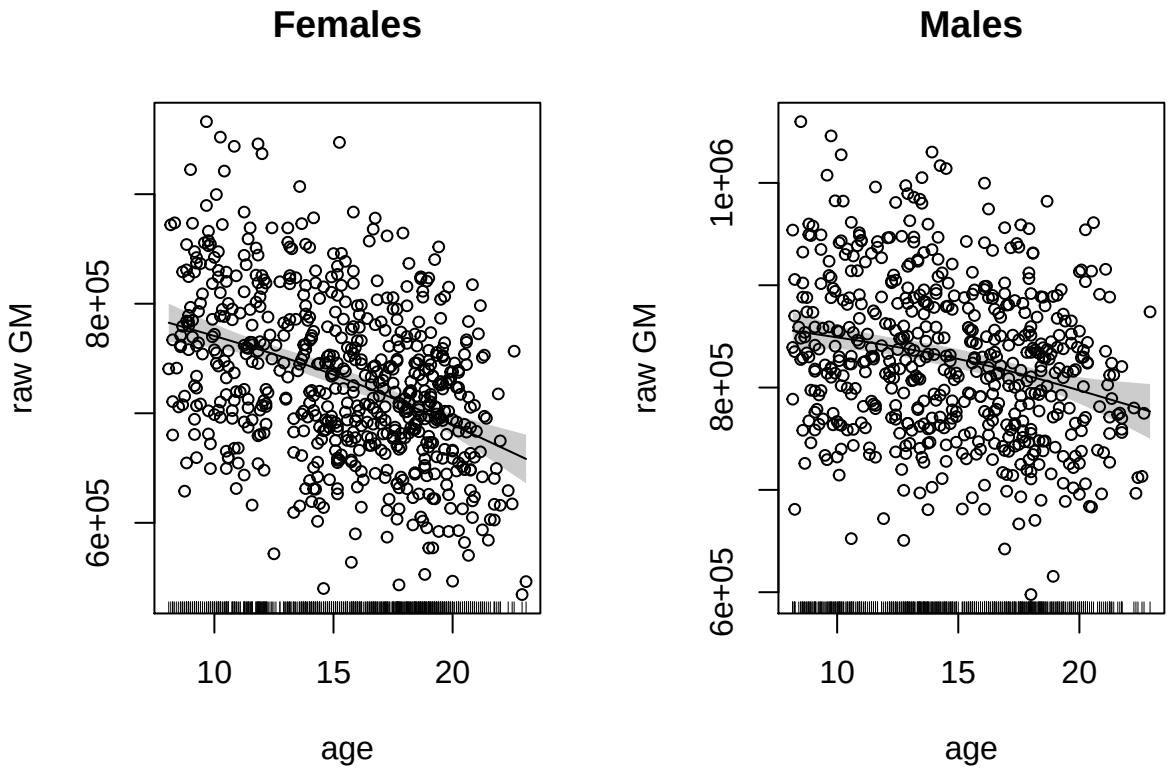
Compare with R-squared of the equivalent linear model (WM ~ sex + age + age x sex): 0.331

Compare GAM and linear model directly:

```
## Analysis of Variance Table
##
## Model 1: WM ~ scale(age) + sex * scale(age) + sex
## Model 2: WM ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##   Res.Df        RSS Df  Sum of Sq      F  Pr(>F)
## 1   1181 2.9777e+12
## 2   1179 2.9615e+12  2 1.6162e+10 3.2172 0.04042 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gray Matter

GM: Plot GAMs



- Males: $R^2 = 0.057$
- Females: $R^2 = 0.16$

GM: Linear models by sex

Table 12:

	Dependent variable:	
	Gray Matter Volume	
	Males	Females
	(1)	(2)
Age	−0.244*** (−0.325, −0.163)	−0.411*** (−0.484, −0.337)
Age (Squared)	−0.032 (−0.113, 0.048)	−0.041 (−0.115, 0.032)
Constant	0.000 (−0.081, 0.081)	−0.000 (−0.072, 0.072)
R^2	0.061	0.163
Note:		*p<0.05; **p<0.01; ***p<0.001
		Stargazer

GM: GAM summary

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## GM ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   731358       2888  253.23   <2e-16 ***
## sexMALE        93289       4223   22.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df     F  p-value
## s(age):sexFEMALE   2      2 55.88  < 2e-16 ***
## s(age):sexMALE     2      2 19.52 4.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.366   Deviance explained = 36.9%
## -REML =  14880  Scale est. = 5.2216e+09  n = 1185
```

Model indicates GM-to-age relationship is linear and statistically significant for both males and females. GM decreases with age.

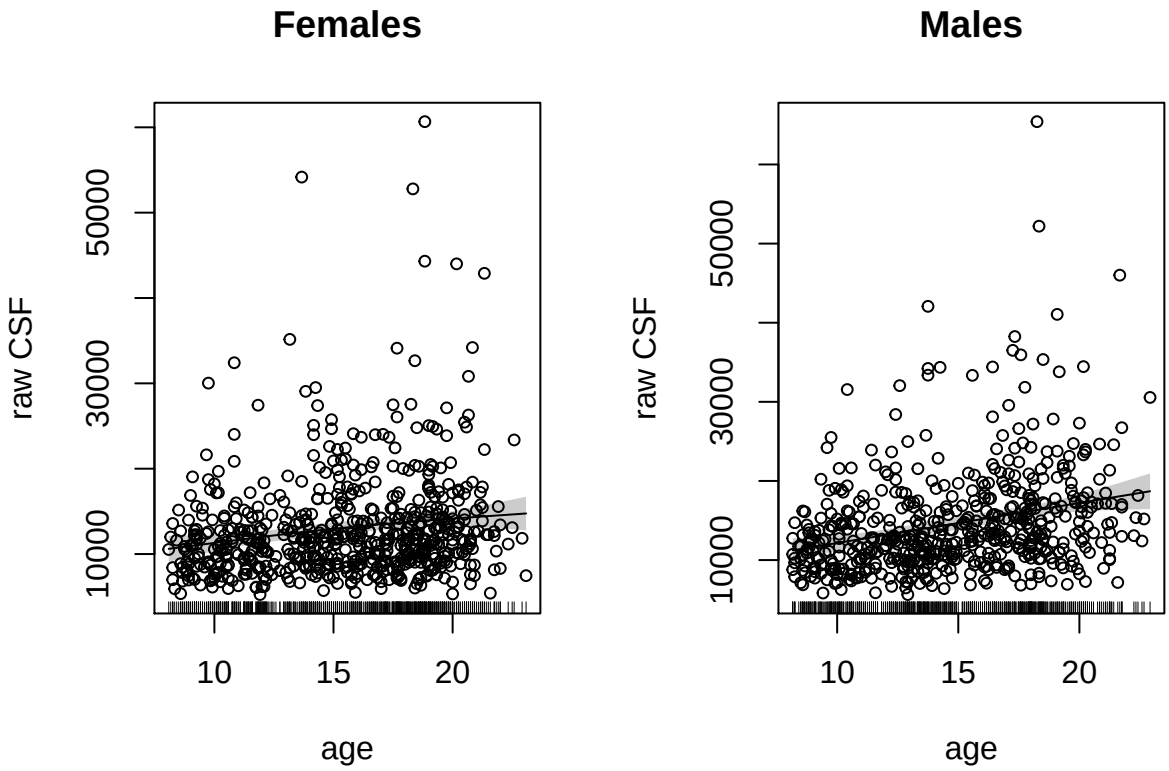
Compare with adj. R-squared of the equivalent linear model (WM ~ sex + age + age x sex): 0.366

Compare GAM and linear model directly:

```
## Analysis of Variance Table
##
## Model 1: GM ~ scale(age) + sex * scale(age) + sex
## Model 2: GM ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##   Res.Df        RSS Df  Sum of Sq      F Pr(>F)
## 1   1181 6.1635e+12
## 2   1179 6.1562e+12  2 7275168233 0.6966 0.4985
```

Cerebrospinal Fluid (in ventricles only)

CSF: Plot GAMs



Compare with R-squared of GAM:
- Males: $R^2 = 0.077$
- Females: $R^2 = 0.026$

CSF: Linear models by sex

Table 13:

	Dependent variable:	
	Cerebrospinal Fluid Volume	
	Males	Females
	(1)	(2)
Age	0.284*** (0.204, 0.364)	0.166*** (0.088, 0.245)
Age (Squared)	−0.002 (−0.082, 0.078)	−0.021 (−0.100, 0.058)
Constant	−0.000 (−0.080, 0.080)	0.000 (−0.077, 0.077)
R^2	0.081	0.030
Note:		*p<0.05; **p<0.01; ***p<0.001
		Stargazer

CSF: GAM summary

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## CSF ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12863.1      249.3  51.589  < 2e-16 ***
## sexMALE       1883.8      364.6   5.167 2.79e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                   edf Ref.df      F  p-value
## s(age):sexFEMALE   2      2  9.239 0.000104 ***
## s(age):sexMALE     2      2 25.171 1.97e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0662   Deviance explained = 7.02%
## -REML =  11992  Scale est. = 3.8917e+07  n = 1185
```

Model indicates CSF-to-age relationship is linear and statistically significant in males and females. CSF increases with age.

Compare with adj. R-squared of the equivalent linear model (CSF ~ sex + age + age x sex): 0.0676

Compare GAM and linear model directly:

```
## Analysis of Variance Table
##
## Model 1: CSF ~ scale(age) + sex * scale(age) + sex
## Model 2: CSF ~ s(age, by = sex, bs = "tp", k = K, fx = FX) + sex
##   Res.Df        RSS Df Sum of Sq     F Pr(>F)
## 1   1181 4.5892e+10
## 2   1179 4.5884e+10  2   8014607 0.103 0.9022
```