An illustration of the relationship between VIF and PCA

As we saw in the previous lecture,

$$VIF_j = \sum_{l=1}^{k} \frac{A_{jl}^2}{L_l},$$

where the A_{jl} were the eigenvector coefficients and the L_l were the eigenvalues from the principal component analysis of the independent variables in a regression model. Here is an illustration of that expression. For the Canadian women's labor-force participation data, the PCA results are:

	W_1	W_2	W_3	W_4	W_5	W_6
Fertility	.385	.668	.542	.252	−.197	−.099
Men's Wages	−.416	.342	−.022	.157	.706	−.432
Women's Wages	−.420	.152	−.266	.729	−.279	.347
Consumer Debt	−.422	.159	−.098	−.276	−.619	−.572
Part-Time Work	−.395	−.469	.775	.152	−.025	−.018
Time	−.411	.411	.158	−.530	.047	.595
Eigenvalue	5.53	.329	.110	.0185	.0071	.0045

For the part-time work variable we have

$$\begin{aligned} VIF_{part\ time\ work} &= \frac{(-.395)^2}{5.53} + \frac{(-.469)^2}{.329} + \frac{(.775)^2}{.110} + \frac{(.152)^2}{.0185} + \frac{(-.025)^2}{.0071} + \frac{(-.018)^2}{.0045} \\ &= .0282 + .669 + 5.46 + 1.25 + .088 + .072 \approx 7.55 \end{aligned}$$

Note how the eigenvector coefficients A_{jl} are small for W_5 and W_6 which gives the part-time work variable a smaller VIF value than for other variables.