Can future crimes be predicted? The first attempt at predictive policing.

Unit 7 Lecture 2

Jonathan Auerbach
STAT 489 Pre-Cap Prof Development
jauerba@gmu.edu

November 11, 2021

How might future crimes be predicted?

These slides use the following R packages

Setup:

```
library("knitr")
library("tidyverse")
theme_set(theme_bw())
```

Can future crimes be predicted?

- France created the first centralized system of crime reporting in 1825.
  - Andre-Michel Guerry (1833), working at the Ministry of Justice, analyzed more than thirty thousand property crimes and ten thousand personal crimes committed between 1825 and 1830.
  - He found the incidence of (reported) crime varied considerably across France. But regular patterns emerged in the data.
    - e.g. crimes against persons consistently highest in summer, crimes against property consistently highest in winter.

- Guerry wondered whether immutable laws—like those describing the phenomena observed in physics—determined crime, ultimately concluding:

  "... the facts of the moral order, like those of the physical order, obey invariant laws, and that, in many respects, the judicial statistics render this a virtual certainty."

Andre-Michel Guerry (1802-1866)

- Guerry was famous in his lifetime, winning the coveted Montyon Prize twice. But he is largely unappreciated today.
  - Friendly (2007) believes Guerry's modesty—both in birth and personality—allowed others to claim credit for his discoveries.

- Nevertheless, his work (along with that of Quetelet) founded the field of "moral statistics" and ultimately sociology and criminology.
  - Additional accomplishments: invented the polar/rose plot, invented a mechanical calculator to compare trends, and was mayor of his village.

Essay on the Moral Statistics of France (1833)

ESSAI

SUR LA

STATISTIQUE MORALE

DE

LA FRANCE,

PRÉCÉDÉ D'UN RAPPORT A L'ACADÉMIE DES SCIENCES,
PAR MM. LACROIX, SILVESTRE ET GIRARD;

PAR A. M. GUERRY,

AVOCAT A LA COUR ROYALE.

Homo, naturæ minister et interpres, de naturæ ordine
tantùm scit et potest quantùm observaverit, nec
amplius scit aut potest.
Bacon, *Nov. Organ. Lib. I, Aph. 1.*

A PARIS,
CHEZ CROCHARD, LIBRAIRE,
RUE ET PLACE DE L'ÉCOLE-DE-MÉDECINE.
LONDRES.
MALIMBOURG, 32 SOUTHAMPTON STR. STRAND, FOREIGN BOOKSELLER.

M DCCC XXXIII.

rien d'arbitraire, et ne tend à favoriser aucun système, puisqu'elle est toute géométrique, et que l'arrondissement de chaque région est déterminé par celui des quatre autres. Elle paraît la plus convenable pour grouper les faits qui doivent être étudiés par grandes masses. Nous sommes loin de desirer, cependant, qu'elle soit adoptée pour la publication des documens de statistique officielle, qui ne sauraient être présentés d'abord avec trop de détails. Si l'administration les donnait seulement, comme on l'a demandé dans la vue de simplifier le travail, par ressort de cour royale ou même pour tous les départemens formant l'arrondissement d'une ancienne province, ils deviendraient bien moins utiles et ne seraient pas recueillis plus aisément.

Si l'on représente par 100 le nombre des crimes commis en France chaque année, les cinq régions offrent les proportions suivantes:

Crimes contre les personnes.

Régions.	Années 1825	1826	1827	1828	1829	1830	Moyenne.
Nord	25	24	23	26	25	24	25
Sud	28	26	22	23	25	23	24
Est	17	21	19	20	19	19	19
Ouest	18	16	21	17	17	16	18
Centre	12	13	15	14	14	18	14
Totaux	100	100	100	100	100	100	100

Crimes contre les propriétés.

Régions.	Années 1825	1826	1827	1828	1829	1830	Moyenne.
Nord	41	42	42	43	44	44	42
Sud	12	11	11	12	12	11	12
Est	18	16	17	16	14	15	16
Ouest	17	19	19	17	17	17	18
Centre	12	12	11	12	13	13	12
Totaux	100	100	100	100	100	100	100

On voit que, pour les crimes contre les personnes, la plus grande différence observée dans chaque région, n'excède jamais de plus de *quatre centièmes*, la moyenne des six années, et que, pour les crimes contre les propriétés, elle n'est pas de plus de *deux centièmes* au-dessus ou au-dessous de cette moyenne. Assurément le produit annuel des récoltes ou des impôts, dans les diverses parties du royaume, ne saurait être évalué d'avance avec plus de précision, de certitude que le nombre des vols, des meurtres et des assassinats.

Sur 100 individus accusés de vol, dans tout le royaume, le nombre des hommes et des femmes a été successivement dans les proportions ci-après:

Guerry's Annual Data (person crimes per thousand)

```
tibble(Year    = 1825:1830,
       North   = c(25, 24, 23, 26, 25, 24),
       South   = c(28, 26, 22, 23, 25, 23),
       East    = c(17, 21, 19, 20, 19, 19),
       West    = c(18, 16, 21, 17, 17, 16),
       Central = c(12, 13, 15, 14, 14, 18)) %>%
kable()
```

Year	North	South	East	West	Central
1825	25	28	17	18	12
1826	24	26	21	16	13
1827	23	22	19	21	15
1828	26	23	20	17	14
1829	25	25	19	17	14
1830	24	23	19	16	18

Guerry's Annual Data (property crimes per thousand)

```
tibble(Year    = 1825:1830,
       North   = c(41, 42, 42, 43, 44, 44),
       South   = c(12, 11, 11, 12, 12, 11),
       East    = c(18, 16, 17, 16, 14, 15),
       West    = c(17, 19, 19, 17, 17, 17),
       Central = c(12, 12, 11, 12, 13, 13)) %>%
kable()
```

Year	North	South	East	West	Central
1825	41	12	18	17	12
1826	42	11	16	19	12
1827	42	11	17	19	11
1828	43	12	16	17	12
1829	44	12	14	17	13
1830	44	11	15	17	13

Essay on the Moral Statistics of France (1833)

INFLUENCE DES SAISONS.

VI.

A. CRIMES CONTRE LES PERSONNES.				B. CRIMES CONTRE LES PROPRIÉTÉS.			
		Sur 1,000.	Sur 1,000.			Sur 1,000.	Sur 1,000.
Hiver.	Décembre.	82	−221	Hiver.	Décembre.	102	279
	Janvier.	69			Janvier.	96	
	Février.	70			Février.	81	
Printemps.	Mars.	85	255	Printemps.	Mars.	84	−236
	Avril.	78			Avril.	75	
	Mai.	92			Mai.	77	
Été.	Juin.	99	+283	Été.	Juin.	78	231
	Juillet.	89			Juillet.	71	
	Août.	95			Août.	82	
Automne.	Septembre.	88	241	Automne.	Septembre.	80	+254
	Octobre.	75			Octobre.	85	
	Novembre.	78			Novembre.	89	
	Totaux.	1,000	1,000		Totaux.	1,000	1,000

Le plus grand nombre des attentats contre les personnes est commis en été; c'est en hiver qu'il y en a le moins. Le printemps et l'automne en présentent un nombre à-peu-près égal (VI, A. — PL. VII, C.).

De tous les crimes contre les personnes, l'attentat à la pudeur est celui pour lequel l'influence des saisons est le plus évidente. Sur 100 crimes de cette espèce, on en compte en été, 36; au printemps, 25; en automne, 21; et en hiver, 18 seulement.

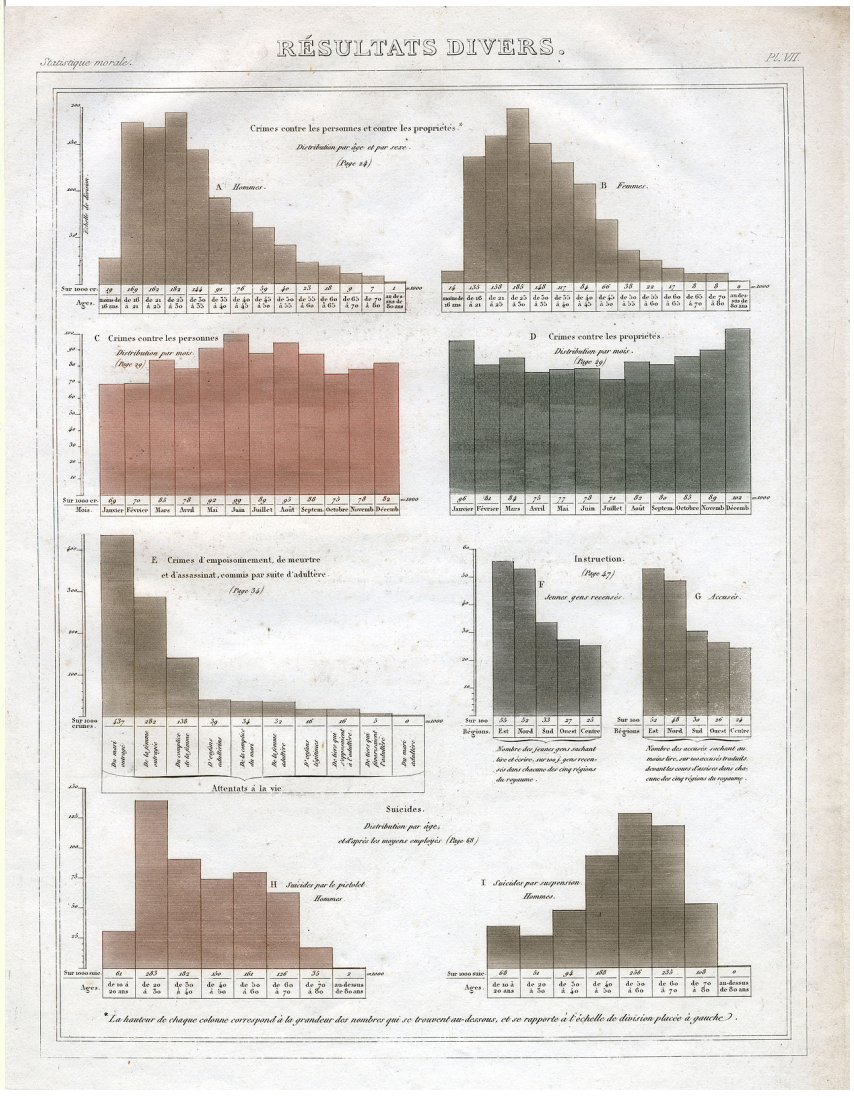
Si cette distribution était, comme on pourrait le croire, un effet indirect des variations de température, les crimes contre les personnes seraient plus nombreux lorsque la température moyenne est le plus élevée, par conséquent, dans les mois de juillet et d'août, tandis que ce n'est pas alors qu'ils le sont ordinairement, mais dans le mois de juin.

Le maximum du nombre des crimes contre les personnes, qui peut néanmoins être lié à l'élévation de la température, paraîtrait coïncider davantage avec la longueur des jours. On pourra s'en assurer dans quelques années, en faisant par mois, des relevés particuliers pour les divers attentats.

L'infanticide est plus fréquent au printemps et en hiver, qu'en été ou en automne. Dans le cas où cette distribution se maintiendrait à l'avenir, elle s'expliquerait aisément, puisque c'est à-peu-près celle des naissances les plus nombreuses. Le mois de mars qui voit commettre le plus d'infanticides est, après celui de février, celui qui compte aussi le plus de naissances.

Les crimes contre les propriétés se présentent à-peu-près en ordre inverse des crimes contre les personnes, de sorte que souvent le minimum des uns coïncide avec le maximum des autres (VI, B. — PL. VII, D.).

C'est en été que les crimes contre les personnes sont le plus fréquens; c'est également pendant cette saison qu'a lieu le plus grand nombre d'admissions dans la maison royale



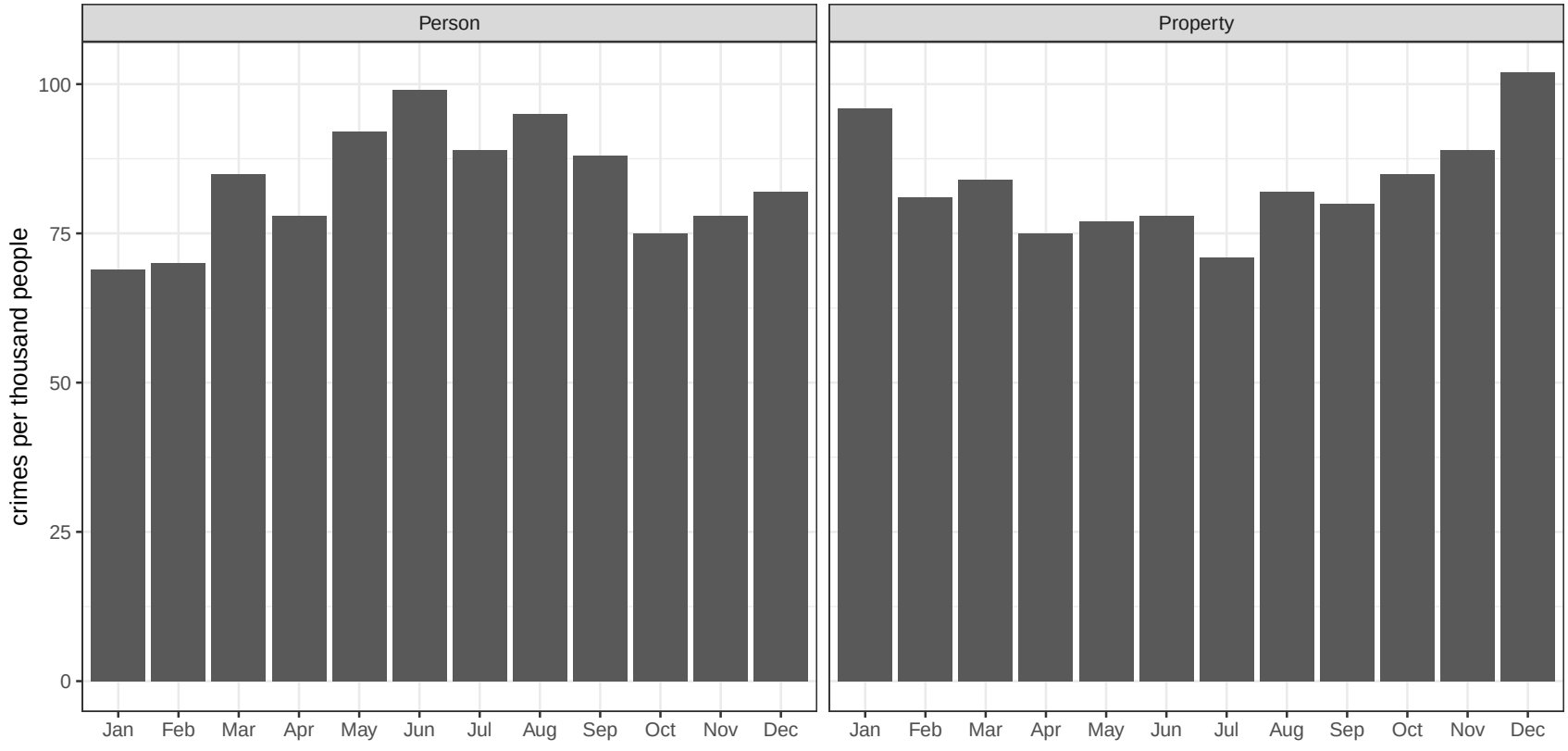
Guerry's Monthly Data (crimes per thousand)

```
Guerry <-
 tibble(Month    = factor(month.abb, levels = month.abb),
        Person   = c(69, 70, 85, 78, 92, 99,
                     89, 95, 88, 75, 78, 82),
        Property = c(96, 81, 84, 75, 77, 78,
                     71, 82, 80, 85, 89, 102))

Guerry %>%
  gather("", Number, -Month) %>%
  spread(Month, Number) %>%
  kable(col.names = c("", str_sub(month.abb, 1, 1)))
```

	J	F	M	A	M	J	J	A	S	O	N	D
Person	69	70	85	78	92	99	89	95	88	75	78	82
Property	96	81	84	75	77	78	71	82	80	85	89	102

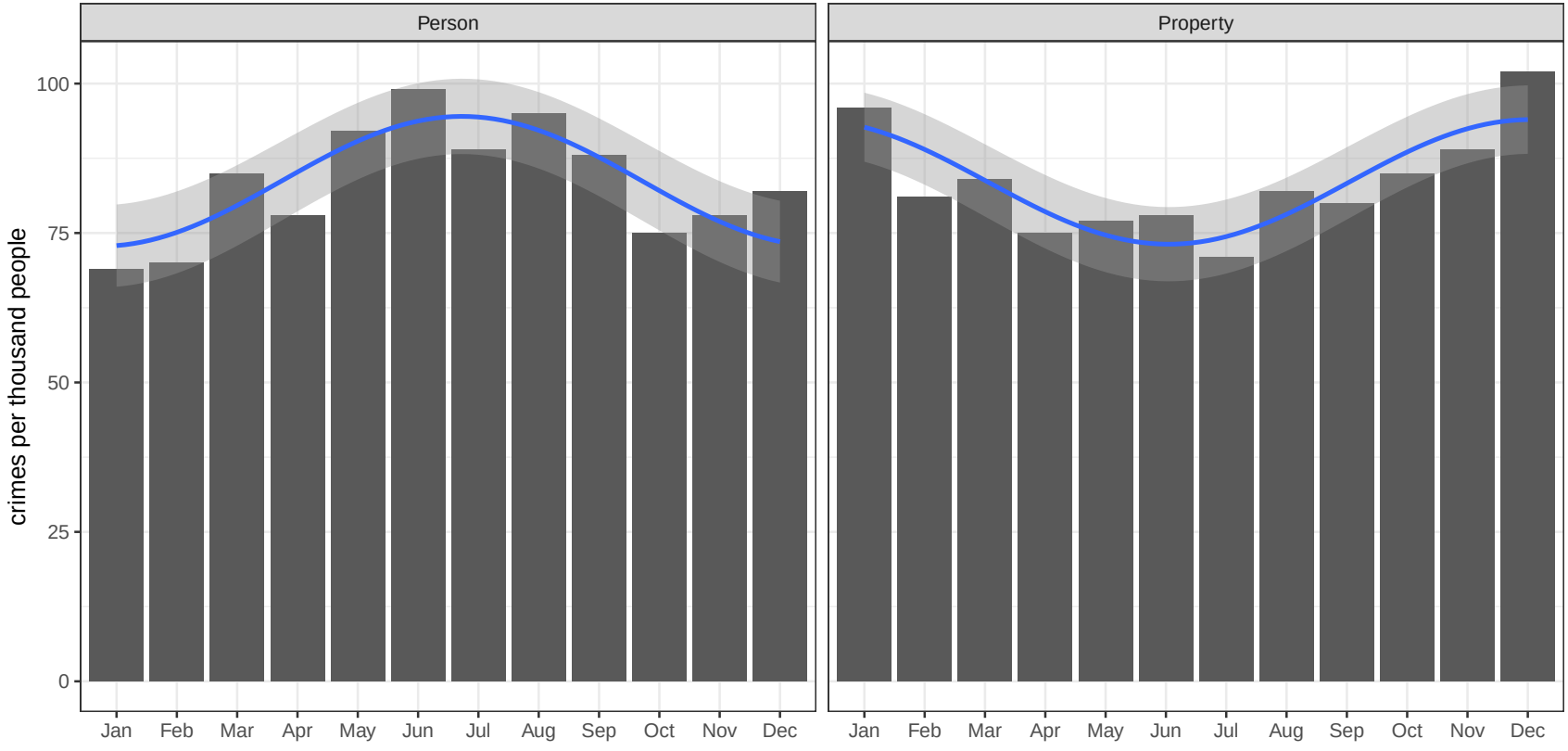
Person crimes greater in summer, property in winter

```
(guerry_plot <-
Guerry %>% gather(type, rate, -Month) %>%
  ggplot(aes(x = Month, weight = rate)) +
    geom_bar() + facet_wrap(~ type) +
    labs(y = "crimes per thousand people", x = ""))
```



Scientists today might describe pattern with sinusoid

```
guerry_plot +
  geom_smooth(aes(as.numeric(Month), rate),
    method = "lm",
    formula = y ~ cos(x*2*pi/12) + sin(x*2*pi/12),
    data = Guerry %>% gather(type, rate, -Month))
```



Harmonic regression: fitting sinusoids with OLS

Suppose $\mathbb{E}[Y|t] = B + A\sin(\omega t + \phi)$ where amplitude A, intercept B, and phase ϕ are unknown. Assume frequency ω is known.

We can estimate A, B, and ϕ by regressing Y on $X_1 = \sin(\omega t)$ and $X_2 = \cos(\omega t)$ since

$$\begin{aligned}
\mathbb{E}[Y|t] &= B + A\sin(\omega t + \phi) \\
&= B + A\cos(\phi)\sin(\omega t) + A\sin(\phi)\cos(\omega t) \qquad \textbf{(1)} \\
&= \beta_1 + \beta_2 \sin(\omega t) + \beta_3 \cos(\omega t) \\
&= \beta_1 + \beta_2 X_1 + \beta_3 X_1
\end{aligned}$$

where **(1)** follows from the Angle Sum Identity:

$$\sin(a + b) = \sin(a)\cos(b) + \cos(a)\sin(b)$$

Note that $A = \sqrt{\beta_2^2 + \beta_3^2}$, $B = \beta_1$, and $\phi = \tan^{-1}(\beta_3/\beta_2)$

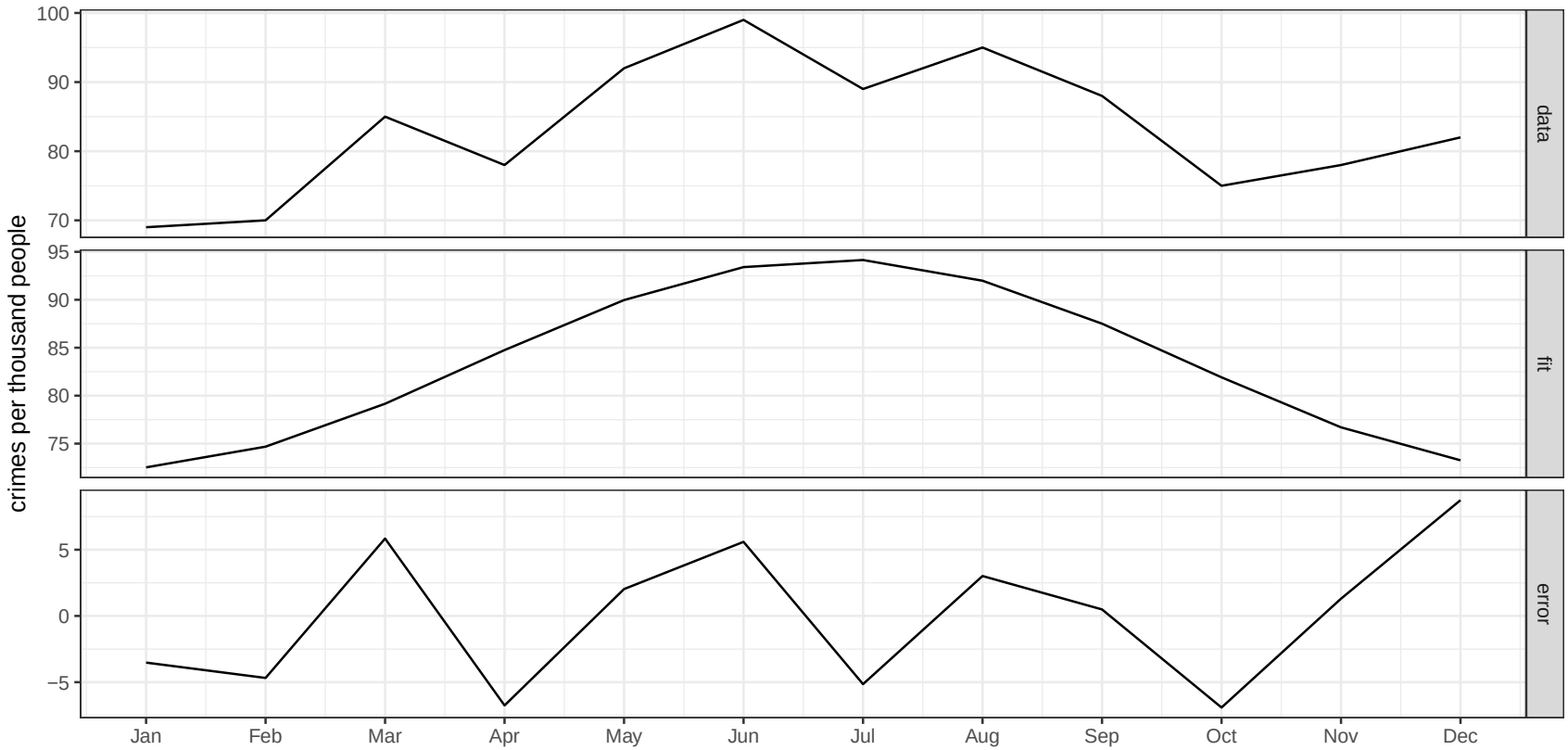
Coefficients for person crimes

```
guerry_fit <- Guerry %>%
  gather(type, rate, -Month) %>%
  filter(type == "Person") %>%
  mutate(x = as.numeric(Month)) %>%
  lm(rate ~ cos(x*2*pi/12) + sin(x*2*pi/12), data = .)

guerry_coef <- guerry_fit %>% coef() %>% unname()

tibble(
 "$\\beta_2$" = guerry_coef[2],
 "$\\beta_3$" = guerry_coef[3],
 "$\\phi$" = atan(guerry_coef[3]/guerry_coef[2]),
 "$A$" = sqrt(guerry_coef[2]^2 + guerry_coef[3]^2)) %>%
  kable(digits = 2)
```

β_2	β_3	ϕ	A
-10.07	-4.18	0.39	10.91

Decomposition of person crimes

```
transmute(Guerry, data = Person, Month,
          fit = predict(guerry_fit), error = data-fit) %>%
  gather(key, value, -Month) %>%
  ggplot() + geom_line(aes(as.numeric(Month), value)) +
  facet_grid(factor(key, lev = c("data","fit","error"))~.,
   scales="free") + labs(y="crimes per thousand people") +
  scale_x_continuous(nam = "", bre = 1:12, lab = month.abb)
```



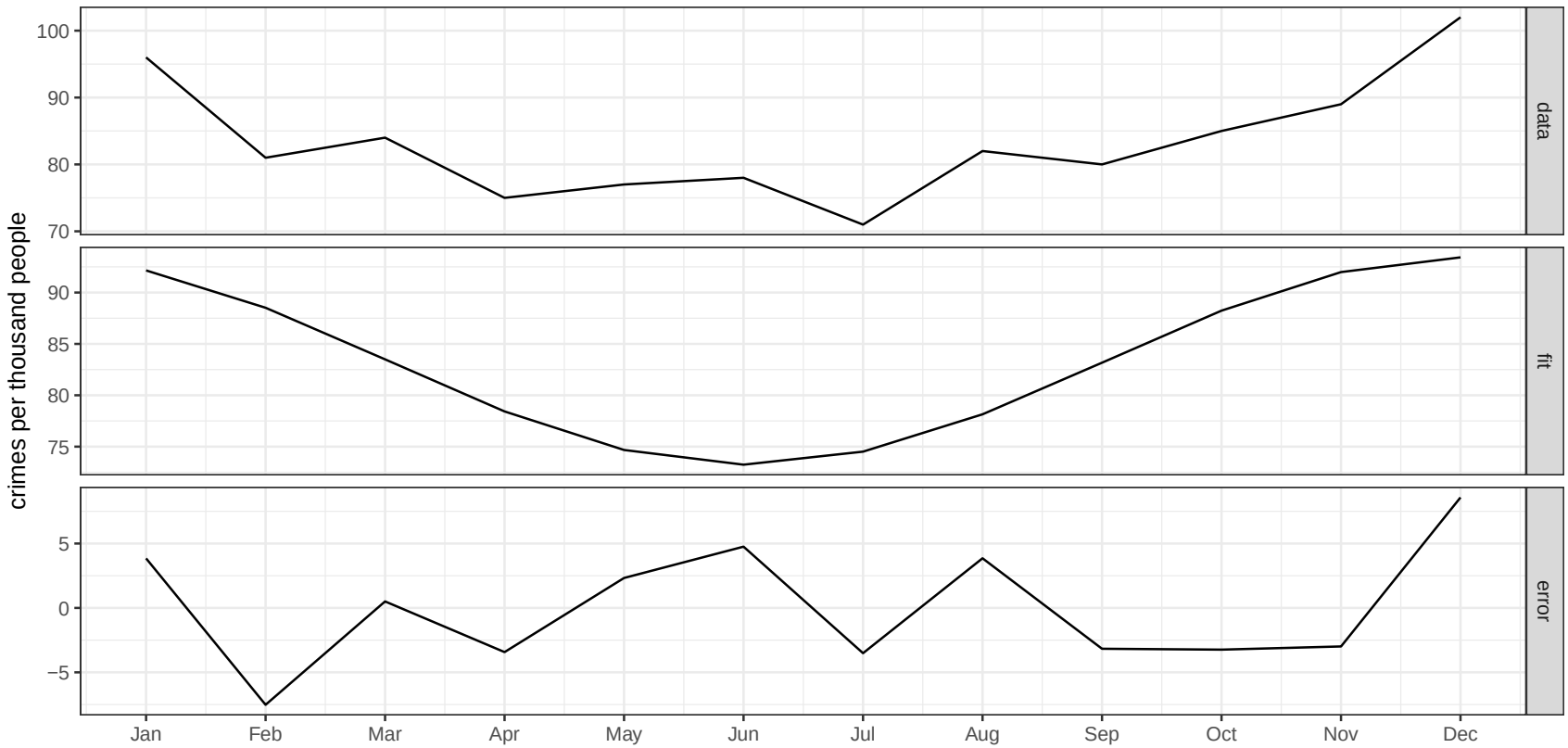
Coefficients for property crimes

```
guerry_fit <- Guerry %>%
  gather(type, rate, -Month) %>%
  filter(type == "Property") %>%
  mutate(x = as.numeric(Month)) %>%
  lm(rate ~ cos(x*2*pi/12) + sin(x*2*pi/12), data = .)

guerry_coef <- guerry_fit %>% coef() %>% unname()

tibble(
 "$\\beta_2$" = guerry_coef[2],
 "$\\beta_3$" = guerry_coef[3],
 "$\\phi$" = atan(guerry_coef[3]/guerry_coef[2]),
 "$A$" = sqrt(guerry_coef[2]^2 + guerry_coef[3]^2)) %>%
  kable(digits = 2)
```

β_2	β_3	ϕ	A
10.09	0.16	0.02	10.09

Decomposition of property crimes

```
transmute(Guerry, data = Property, Month,
          fit = predict(guerry_fit), error = data-fit) %>%
  gather(key, value, -Month) %>%
  ggplot() + geom_line(aes(as.numeric(Month), value)) +
  facet_grid(factor(key, lev = c("data","fit","error"))~.,
   scales="free") + labs(y="crimes per thousand people") +
  scale_x_continuous(nam = "", bre = 1:12, lab = month.abb)
```



Are predicted crimes preventable crimes?

- Governments have used data to inform policing to various degrees since Guerry; "the test of police efficiency is the absence of crime and disorder, and not the visible evidence of police action in dealing with them." (Principles of Policing, Peel 1829)
  - The current era of data-driven policing began when NYPD created its real-time crime reporting system, CompStat (Compare Stats or Computer Statistics portmanteau, 1994).
  - NYPD uses CompStat to predict and then target areas for specific crime prevention strategies.
    - Between 1990 and 2009, homicide, robbery, and burglary fell over 80 percent. Drug-related violence fell more than 90 percent.
    - Moreover, incarceration did not rise dramatically, challenging the previous zero-tolerance paradigm.

- Yet, these tactics were still controversial.
  - In 2013, a judge ruled the NYPD's widespread practice of "stop, question and frisk" allowed for searches that violated the US Constitution.
  - Critics also argued the dependence on data incentivized quotas and false reporting.

Post hoc ergo propter hoc

- It is unclear whether NYPD strategies caused the crime reduction; whether crime would remain high if the old strategy had been continued.
  - Claiming crime fell because it followed a change in strategy is the "post hoc ergo propter hoc" fallacy.
  - Just because one event proceeds another does not mean the first is necessarily the cause of the second.

- One event is said to Granger cause another if:
  1. the cause occurs before the effect, and
  2. the cause contains information about the effect not available from another source (Clive Granger, 1969; Nobel Prize 2003).

- Major socioeconomic changes occurred in NYC between 1990 and 2009, such as immigration, gentrification, and the rise of the personal computer. These changes may better explain the drop in crime.

References

1. Friendly, Michael. "The Life and Works of André-Michel Guerry (1802-1866)." (2007)

2. Granger, Clive WJ. "Time series analysis, cointegration, and applications." American Economic Review 94.3 (2004): 421-425.

3. Guerry, A. M. "Statistique morale de l´ Angleterre comparée avec la statistique morale de la France." (1864).

4. Guerry, A. M. "A Translation of Andre-Michel Guerry's Essay on the Moral Statistics of France." (2002).

5. Zimring, Franklin E. "The city that became safe." New York's lessons for urban crime and (2012).