Kevin Wang PhD

Data Scientist and Statistician

(+61) 404 955 255
kevinwangstats@gmail.com
Portfolio: kevinwang09.github.io
Nationality: Australian

Experience

Present 2020 Jul **Statistician (data scientist)**, *CSL Behring*, Australia.

CSL is the largest biotechnology company in Australia, manufacturing plasma protein derivatives and vaccines. Selected work responsibilities:

- Using statistical process control techniques to standardise manufacturing and quality control laboratory procedures with engineers and biochemists.
- Using innovative machine learning models and R-Shiny dashboards to identify areas of improvement, automate routine procedures and speed up turnaround time for QC and manufacturing.
- Biostatistical support for R&D of new products, including high-value products like coronavirus vaccines and treatments.
- Design experiments and methodologies to understand the chemical stability of products, conform to regulatory requirements and ensure product effectiveness.
- Achieved savings of multiple batches of plasma products, with some batch valued up to $1M.

2020 Jun 2019 Sep **Research Associate**, *University of Sydney*.

- The major research project developed a framework for clinical outcome prediction for individual patients using genomics data. Melanoma was a key focus of this research and the key biomarkers were validated and implemented on a cost-effective bioassay.
- `R` software packages associated with the research were developed and maintained in a reproducible environment.
- Visiting scholar at Cornell University for 2 months and delivered a workshop in open-source single-cell analytics using Cloud technology.
- Published peer-reviewed articles in top-tier journals covering a wide range of topics in cancer and public health.

2019 Mar 2016 Mar **Postgraduate Teaching Fellow**, *University of Sydney*.

- Delivered lectures (200+ students) and tutorials, covering 15 different courses, including statistics, mathematics and data science. Achieved 85% satisfactory rating for end-of-course surveys.
- Authored and designed course content at both the undergraduate and postgraduate levels.
- Mentored and trained new statistics tutors.

Education

2016–2020 **Doctor of Philosophy in Science**, *University of Sydney*.

- The PhD research was on statistical bioinformatics.
- Statistical machine learning methods were developed to solve specific challenges in individual patient clinical outcome predictions and thereby enabling personalised medicine.
- Real-world biomedical data from collaborators were used to motivate and articulate important challenges in current statistical research. The value of this thesis is the synthesis of statistics, clinical implementation and practical deployment.

2012-2015 **Bachelor of Science (Adv. Mathematics) (Hon. I)**, *University of Sydney*.

- Major in statistics and financial mathematics.
- The Honours thesis examined functional MRI data and the inferred brain connectivity networks through the use of different statistical association measures.

Skills

Applied statistics	Highly experienced at the research level and consulting level. Ability to transform business and cross-disciplinary problems into a statistical framework to deliver solutions. Strong experience with model selection to extract explainable insights from data.
Statistical modelling	Highly experienced at the research level and consulting level. Specialisation in predictive modelling for high-dimensional data such as genomics data with 20k+ features.
`R` & `git`	Highly proficient in scripting under a reproducible environment. Strong ability to use `tidyverse` and `shiny` to deal with dirty data and produce visualisations and presentable insights. Ability to combine git together with unit testing, continuous integration and coverage testing. Author and maintainer of several open-source `R` packages, including `APES`, `mcvis` and `scMerge`.
`python` & `SQL`	Proficient in writing scripts and notebooks. Experience in `pandas`, `numpy`, `scikit-learn` and `keras` for personal interest projects on GitHub.
docker & Cloud	Proficient in building docker images and deployment through Google Cloud. Experience in holding reproducible workshops using both with excellent participant feedback.

Volunteering

Present – 2021 **Communication officer**, *Statistical Society of Australia, Victorian branch*.
Managing communications and social media accounts. Organising events and workshop for the society.

2017 **President**, *Sydney University Mathematics Society*.
Oversaw finances of the Society, organised academic/industry events and negotiated corporate sponsorships for three years before the presidency in 2017. Delivered seminars and hands-on workshops for visiting high school and university students.

Scholarships & Awards

2019 **Statistical Society of Australia, NSW**, *JB Douglas Award (joint runner-up). Competitive prize awarded to the top statistics research PhD candidates in New South Wales*.

2019 **Statistical Society of Australia**, *Golden Jubilee Travel Grant. Competitive funding for a statistics PhD candidate in Australia*.

2017 **International Biometric Society - Australasian Region**, *Best student talk at the biennial conference*.

2016-2020 **Australian Postgraduate Award**, *For the duration of the PhD program at the University of Sydney*.

2016 **Charles Perkins Centre**, *Summer Research Scholar in Bioinformatics. Research in data visualisation, focus on gene interaction networks for cancers*.

2015 **International Biometric Society - Australasian Region**, *Awarded to the top biostatistics student undertaking a Honours program in AU/NZ region*.

2014 **Australian National University**, *Summer Research Scholar in Mathematical Statistics. Research in statistical model selection and averaging techniques*.

2013 **Winston Charitable Foundation**, *Research Internship in Mathematical Biology. Research in mathematical modelling in the spread of infectious parasites in beehives*.

Publications

1 Schafer, S., **Wang, K. Y. X.**, Sundling, F., Yang, J.Y.H., & Liu, A. **2021**. Modelling maternal and perinatal risk factors to predict poorly controlled childhood asthma. *PLOS ONE*, (in press).

2 Kim J.H., **Wang, K. Y. X.**, Chen, C., Lin, Y., Tam, P.P.L., Lin, D.M., Yang, J.Y.H., & Yang, P. **2021**. Cepo uncovers cell identity through differential stability. (Under review).

3 Lin, C., **Wang, K. Y. X.**, & Mueller, S. **2020**. mcvis: A new framework for collinearity discovery, diagnostic and visualization. *Journal of Computational and Graphical Statistics*, 1-13.

4 Hewavisenti, R., Ferguson, A., **Wang, K. Y. X.**, Jones, D., Gebhardt, T., Edwards, J., Zhang, M., Britton, W., Yang, J., Hong, A., & Palendira, U. **2020**. CD103+ tumour-resident CD8+ T cell numbers underlie improved patient survival in oropharyngeal squamous cell carcinoma. *Journal for ImmunoTherapy of Cancer*, 8:e000452.

5 **Wang, K.Y.X.**, Tarr, G., Yang, J.Y.H., Mueller, S. **2019**. Fast and approximate exhaustive variable selection for generalised linear models with APES, Invited paper to *Australia & New Zealand Journal of Statistics*, 61 (4) 445-465.

6 Lin, Y., Ghazanfar, S., **Wang, K.Y.X.**, Gagnon-bartsch, J.A., Lo, K.K., Han, Z., Ormerod, J.T., Speed, T.P., Yang, P., Yang, J.Y.H. **2019**. scMerge: Leveraging factor analysis, stable expression and pseudo-replication to merge multiple single-cell RNA-seq data, *Proceedings of the National Academy of Sciences of the United States of America*, 116 (20) 9775-9784.

7 Pires da Silva, I., **Wang, K.Y.X.**, Wilmott, J.S., Holst, J., Carlino, M.S., Park, J.J., Quek, C., Wongchenko, M., Yan, Y., Mann, G., Johnson, D.B., McQuade, J.L., Rai, R., Kefford, R.F., Rizos, H., Scolyer, R.A., Yang, J.Y.H., Long, G. V, Menzies, A.M. **2019**. Distinct molecular profiles and immunotherapy treatment outcomes of V600E and V600K BRAF-mutant melanoma. *Clinical Cancer Research*, 25 (4) 1272-1279.

8 **Wang, K.Y.X**., Menzies, A.M., Silva, I.P., Wilmott, J.S., Yan, Y., Wongchenko, M., Kefford, R.F., Scolyer, R.A., Long, G. V, Tarr, G., Mueller, S., Yang, J.Y.H. **2019**. bcGST - an interactive bias-correction method to identify over-represented gene-sets in boutique arrays. *Bioinformatics*, 35 (8) 1350-1357.

9 Strbenac, D., **Wang, K.Y.X.**, Wang, X., Dong, J., Mann, G.J., Mueller, S., Yang, J.Y.H. **2019**. Melanoma Explorer: a web application to allow easy reanalysis of publicly available and clinically-annotated melanoma omics datasets. *Melanoma Research*.

Seminars

Workshop
- *"Single-cell RNA-Seq analysis from beginning to end"* at Hong Kong University (2019), Sydney University (2019) and Cornell University (2020).
- *"Enter the Tidyverse with R and RStudio"* at BioinfoSummer (2019).
- *"Fast algorithms and modern visualisations for feature selection"* at Joint International Society for Clinical Biostatistics and Australian Statistical Conference (2018).

Talks
- 2019 International Conference on Econometrics and Statistics (invited)
- 2019, 2017 International Biometrics Society - Australasian Region Conference
- 2018 Australian Bioinformatics & Computational Biology Society Conference
- 2018 Joint Statistical Meeting
- 2017 International Conference in Robust Statistics
- 2016 Australian Statistical Conference